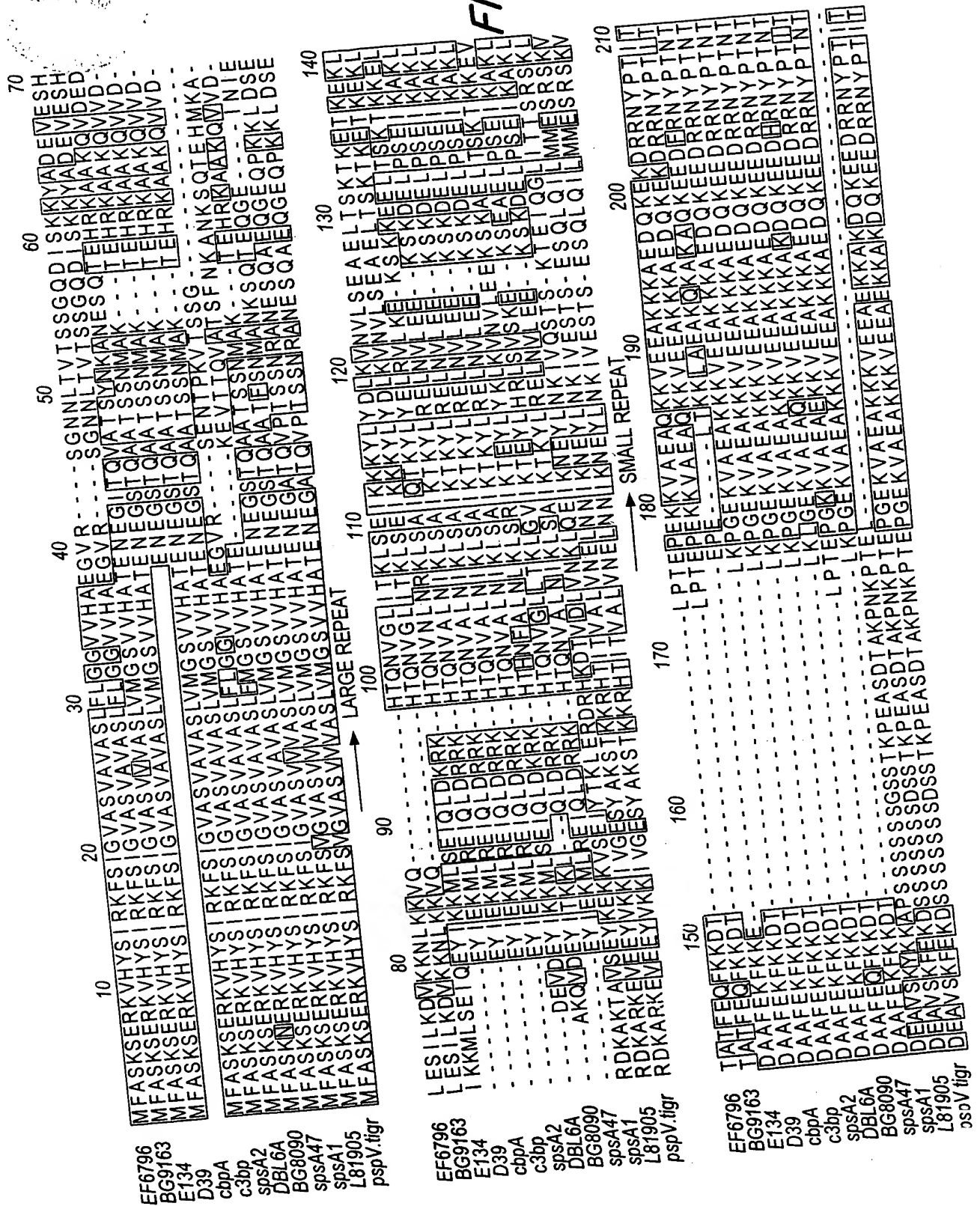
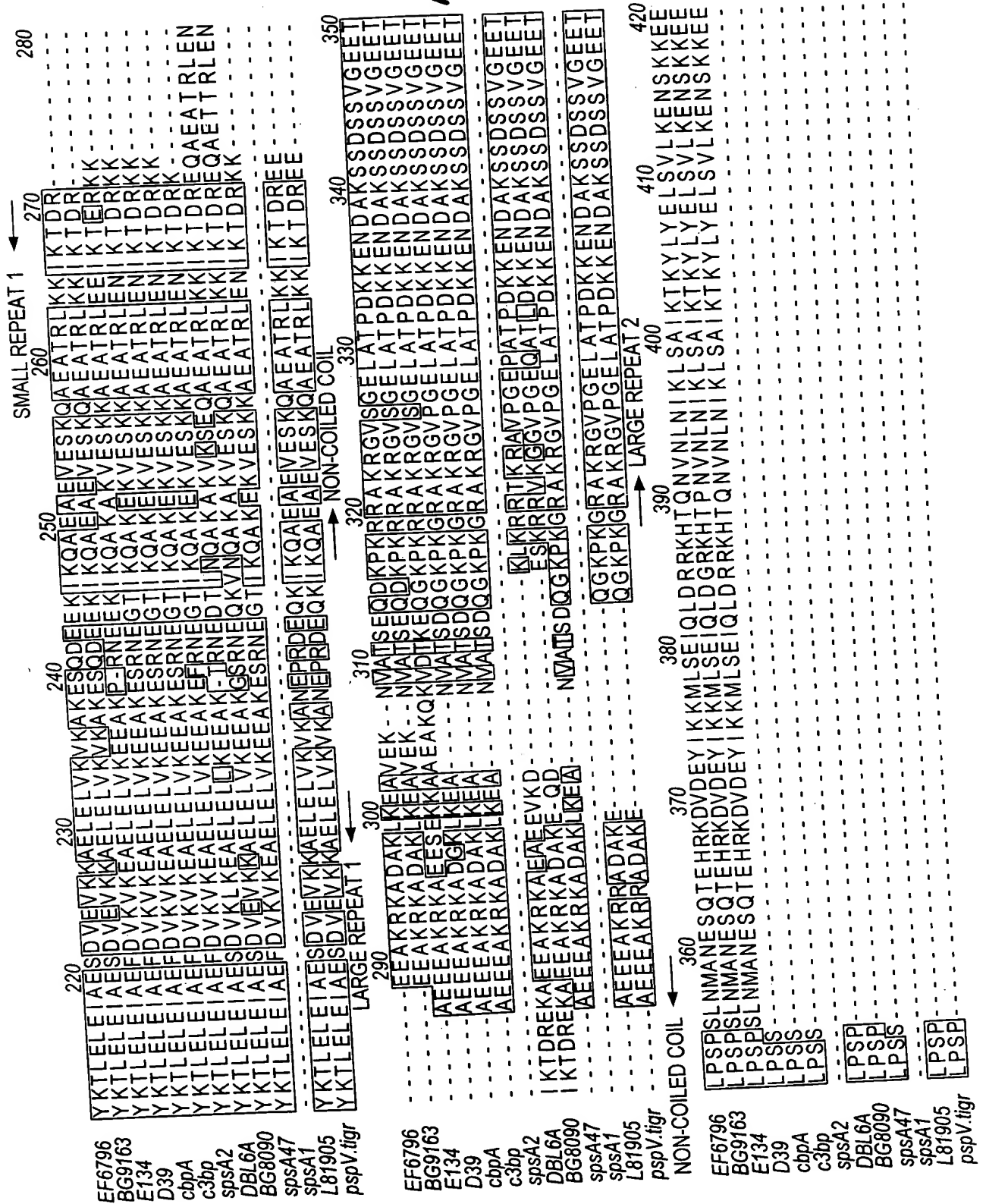
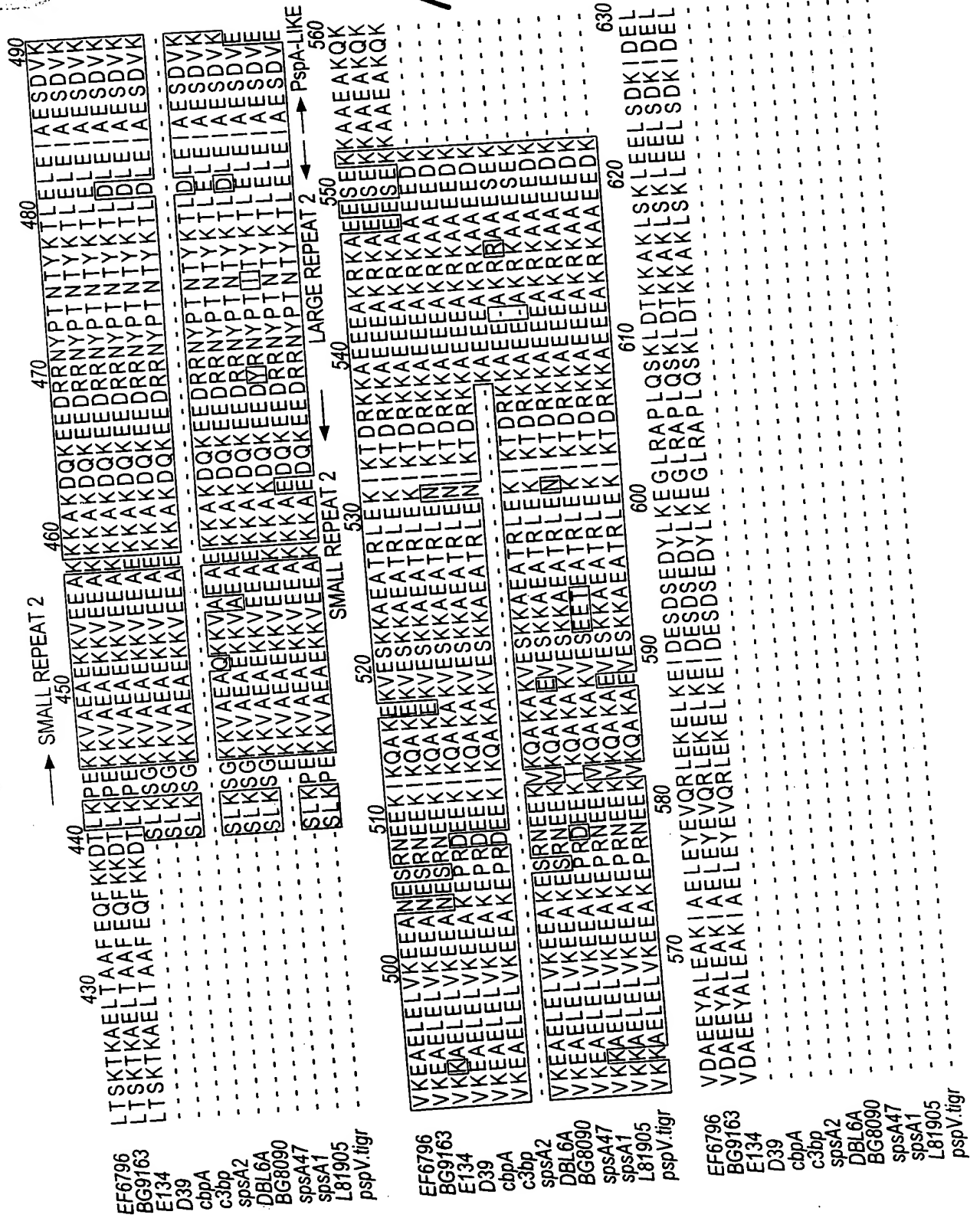


FIGURE 1: CARTOON OF THE PspC CLADES COMPARED TO A REPRESENTATIVE PspA MOLECULE. LONG ARROWS REPRESENT THE DIRECT REPEATS FOUND WITHIN  $\alpha$  HELIX. THE HYPERVARIABLE REGION IS INDICATED BY THE ZIG-ZAG LINES. THE REGION OF HOMOLOGY FOUND WITH THE ALPHA HELIX IS INDICATED BY THE HORIZONTAL LINES.

**FIG. 1**







**FIG. 2D**

	a	b	c	d	e	f	g	
1								
11								Glu Gly Val Arg Ser Gly Asn Asn Leu Thr
16								Val Thr Ser Ser Gly
22	Gln	Asp	Ile	Ser	Lys	Lys		
29	Tyr	Ala	Asp	Glu	Val	Glu	Ser	
34		His	Leu	Glu	Ser	Ile		
41	Leu	Lys	Asp	Val	Lys	Lys	Asn	
44	Leu	Lys	Lys					
51	Val	Gln	His	Thr	Gln	Asn	Val	
56		Gly	Leu	Ile	Thr	Lys		
63	Leu	Ser	Glu	Ile	Lys	Lys	Lys	
64							Tyr	
69	Leu	Tyr	Asp	Leu	Lys			
76	Val	Asn	Val	Leu	Ser	Glu	Ala	
81		Glu	Leu	Thr	Ser	Lys		
85		Thr	Lys	Glu	Thr			
92	Lys	Glu	Lys	Leu	Thr	Ala	Thr	
99	Phe	Glu	Gln	Phe	Lys	Lys	Asp	
105								Thr Leu Pro Thr Glu Pro
108					Glu	Lys	Lys	
115	Val	Ala	Glu	Ala	Gln	Lys	Lys	
122	Val	Glu	Glu	Ala	Lys	Lys	Lys	
126					Ala	Glu	Asp	Gln
133	Lys	Glu	Lys	Asp	Arg	Arg	Asn	
138	Tyr	Pro	Thr	Ile	Thr			
145	Tyr	Lys	Thr	Leu	Glu	Leu	Glu	
152	Ile	Ala	Glu	Ser	Asp	Val	Glu	
159	Val	Lys	Lys	Ala	Glu	Leu	Glu	
166	Leu	Val	Lys	Val	Lys	Ala	Lys	
173	Glu	Ser	Gln	Asp	Glu	Glu	Lys	
180	Ile	Lys	Gln	Ala	Glu	Ala	Glu	
187	Val	Glu	Ser	Lys	Gln	Ala	Glu	
190					Ala	Thr	Arg	
197	Leu	Lys	Lys	Ile	Lys	Thr	Asp	
204	Arg	Glu	Glu	Ala	Lys	Arg	Lys	
210		Ala	Asp	Ala	Lys	Leu	Lys	
216		Glu	Ala	Val	Glu	Lys	Asn	
223	Val	Ala	Thr	Ser	Glu	Gln	Asp	
224	Lys							
234								Pro Lys Arg Arg Ala Lys Arg Gly Val Ser
244								Gly Glu Leu Ala Thr Pro Asp Lys Lys Glu
254								Asn Asp Ala Lys Ser Ser Asp Ser Ser Val
264								Gly Glu Glu Thr Leu Pro Ser Pro Ser Leu
268								Asn Met Ala Asn
271					Glu	Ser	Gln	
277	Thr	Glu	His	Arg	Lys	Asp		
281	Val	Asp	Glu	Tyr				
288	Ile	Lys	Lys	Met	Leu	Ser	Glu	
295	Ile	Gln	Leu	Asp	Arg	Arg	Lys	
300		His	Thr	Gln	Asn	Val		
305		Asn	Leu	Asn	Ile	Lys		
312	Leu	Ser	Ala	Ile	Lys	Thr	Lys	
316					Tyr	Leu	Tyr	Glu
323	Leu	Ser	Val	Leu	Lys	Glu	Asn	
325							Ser	Lys
332	Lys	Glu	Glu	Leu	Thr	Ser	Lys	
336					Thr	Lys	Ala	Glu
	Leu	Thr	Ala	Ala	Phe	Glu	Gln	

FIG.3A

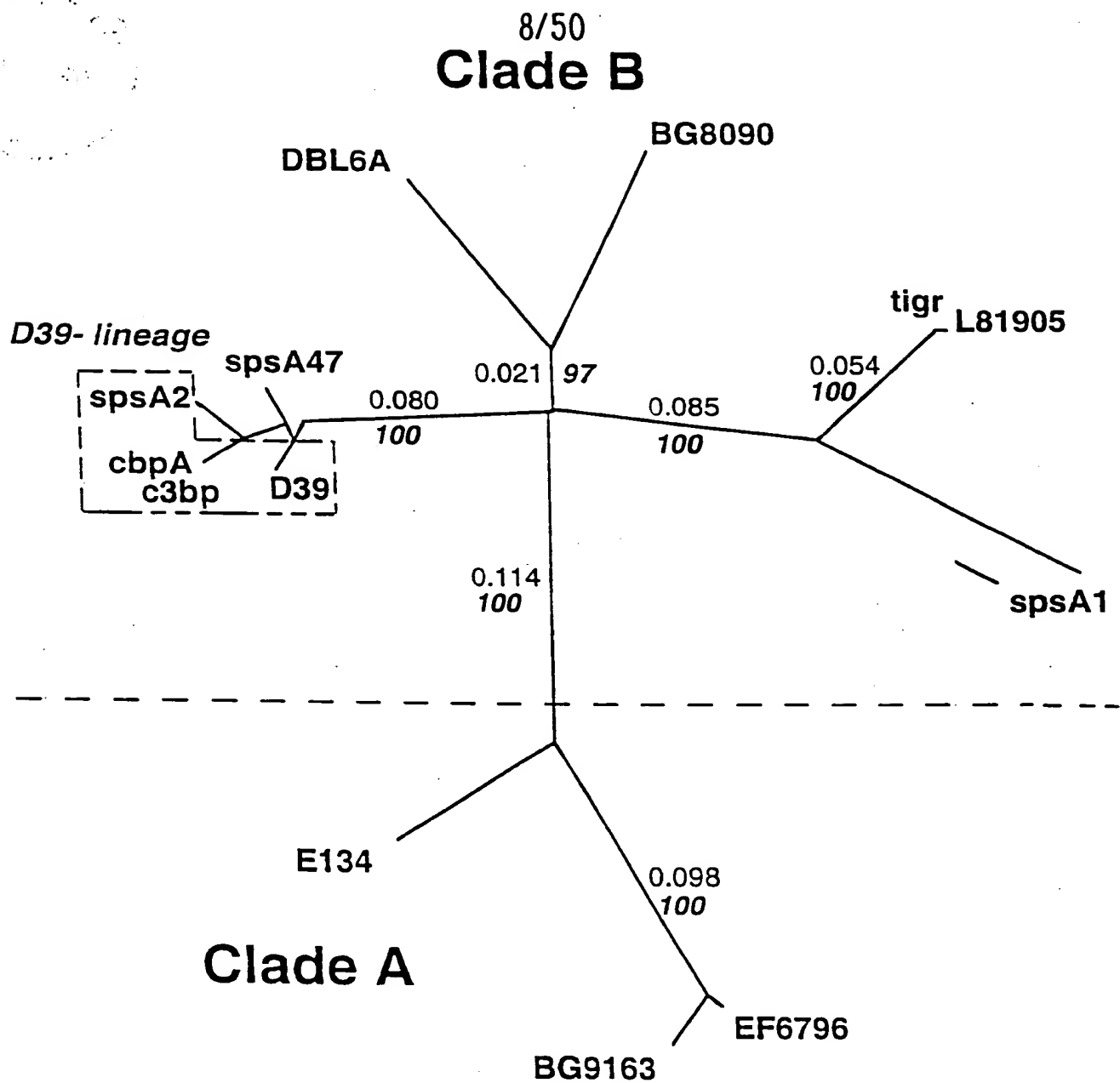
```

343 Phe Lys Lys
346
351
354 Val Ala Glu Ala Glu Lys Lys
361 Val Glu Glu Ala Lys Lys Lys
368 Ala Lys Asp Gln
372 Lys Glu Glu Asp Arg Arg Asn
379 Tyr
380 Pro Thr Asn Thr
384 Tyr Lys Thr Leu Glu Leu Glu
391 Ile Ala Glu Ser Asp Val Lys
398 Val Lys Glu Ala Glu
403 Leu Glu Leu Val Lys Glu Glu
410 Ala Asn Glu Ser Arg Asn Glu
417 Glu Lys Ile Lys Gln Ala
423 Lys Glu Lys Val Glu Ser Lys
430 Lys Ala Glu Ala Thr Arg
436 Leu Glu Lys Ile Lys Thr Asp
443 Arg Lys Lys Ala Glu Glu Glu
450 Ala Lys Arg Lys
454 Ala Glu Glu Ser Glu Lys Lys
461 Ala Ala Glu Ala Lys Gln Lys
468 Val Asp Ala Glu Glu Tyr Ala
475 Leu Glu Ala Lys
479 Ile Ala Glu Leu Glu Tyr Glu
486 Val Gln Arg Leu Glu Lys Glu
493 Leu Lys Glu
496 Ile Asp Glu Ser Asp Ser Glu
503 Asp Tyr Leu Lys Glu Gly
509 Leu Arg Ala
512 Pro Leu Gln Ser Lys
517 Leu Asp Thr Lys Lys Ala Lys
524 Leu Ser Lys
527 Leu Glu Glu Leu Ser Asp Lys
534 Ile Asp Glu Leu Asp Ala Glu
541 Ile Ala Lys Leu Glu Val Gln
548 Leu Lys Asp Ala Glu Gly Asn
555 Asn Asn
557 Val Glu Ala Tyr Phe Lys Glu
564 Gly Leu Glu Lys Thr
569 Thr Ala Glu Lys Lys
574 Ala Glu Leu Glu Lys Ala
580 Glu Ala Asp Leu Lys Lys Ala
587 Val Asp Glu

```

The coiled-coil motif of the  $\alpha$ -helix of PspC. Amino acids that are not in the coiled-coil motif are in the right column. This is the output from the Matcher program accessed through the Internet (<http://catt.poly.edu/~jps/>).

**FIG.3B**



————— 0.05 changes

**FIG.4**



## Choline-binding Domain

PspC consensus

11111111112

12345678901234567890

- 1 KTGWKQENGWYFYNTDGSMA
- 2 TGWLQNNGSWYYLNAMGAMA
- 3 TGWLQNNGSWYYLNANGSMA
- 4 TGWLQNNGSWYYLNANGAMA
- 5 TGWLQYNGSWYYLNANGDMA
- 6 TGWLQYNGSWYYLNSNGAMA
- 7 TGWLQYNGSWYYLNANGDMA
- 8 TGWLQNNGSWYYLNANGDMA
- 9 TGWLQYN SWYYLNANGDMA
- 10 TGWVKDGDWYYLEASGAMKA
- 11 SQWFKVSDKWYYVNGSGALA

VNTTVDGYGVNANGWVN

18 amino acid tail PspC  
28% identical to consensus

TGWLQNNGSWYYLNANGAMA PspC consensus repeat

Y

For PspA/R36A alignment see Yother and Briles 1992.

TGWLQXNGSWYYLNANGAMA

Y

PspA/Rx1 consensus  
95% identical to PspC consensus

VNTTVDGYKVNANGEWV\_

PspA/Rx1 17 amino acid tail.  
One AA different from PspC,  
And one AA shorter.

FIG.5A

10/50

PspA/EF5668

11111111112

12345678901234567890

1 IGWKQENG MWYFYNTDGSM A  
2 TGWLQNNGSWYYLNSNGAM A  
3 TGWLQYNGSWYYLNANGAM A  
4 TGWLQYNGSWYYLNANGAM A  
5 TGWLQYNGSWYYLNANGDM A  
6 TGWLQYNGSWYYLNANGDM A  
7 TGWAKVHGSWYYLNANGSM A  
8 TGWVKDGETWYYLEASGSMKA  
9 NQWFQVSDKWYYVNGLGSL S  
10 VNTTVDGYKVNANGEWV

17 amino acid tail EF5668 PspA

TGWLQYNGSWYYLNANGSM A

PspA/EF5668 consensus

90% identical to PspC consensus

Repeat #1

PspC consensus

KTGWKQENG MWYFYNTDGSM A

PspA/Rx1

TGWKQENG MWYFYNTDGSM A

PspA/EF5668

IGWKQENG MWYFYNTDGSM A

Repeat #N-1

PspC consensus

TGWVKDGD TWYYLEASGAMKA

PspA/Rx1

TGWVKDGD TWYYLEASGAMKA

PspA/EF5668

TGWVKDGD TWYYLEASGSMKA

Repeat #N

PspC consensus

SQWFKVSDKWYYVNGSGALA

PspA/Rx1

SQWFKVSDKWYYVNGLGALA

PspA/EF5668

NQWFQVSDKWYYVNGLGSL S

17-18 AA tail

PspC consensus

VNTTVDGYGVNANGEWV

PspA/Rx1

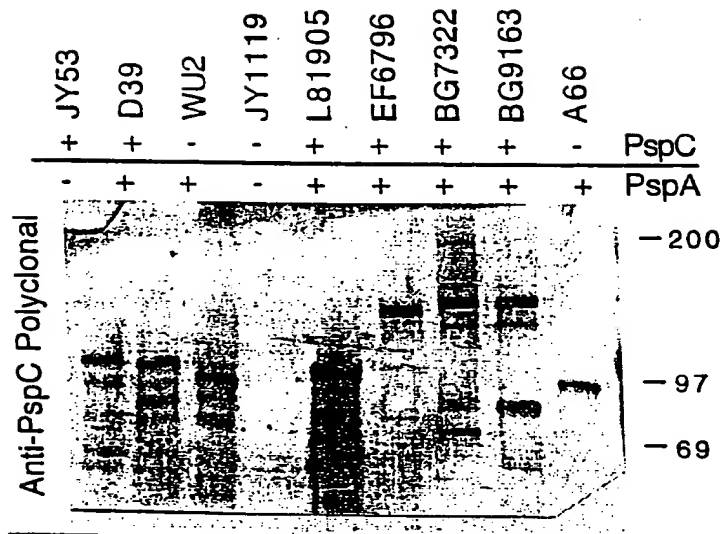
VNTTVDGYKVNANGEWV

PspA/EF5668

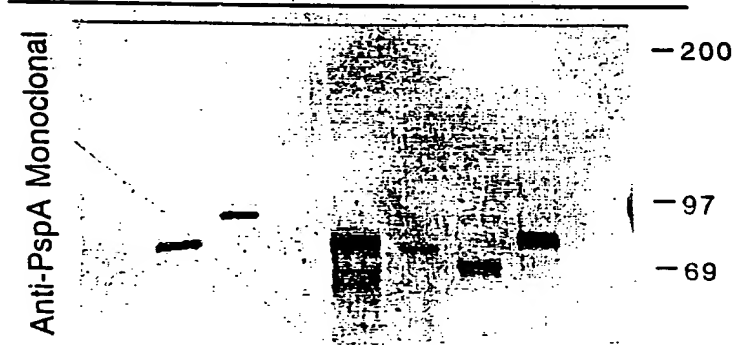
VNTTVDGYKVNANGEWV

**FIG.5B**

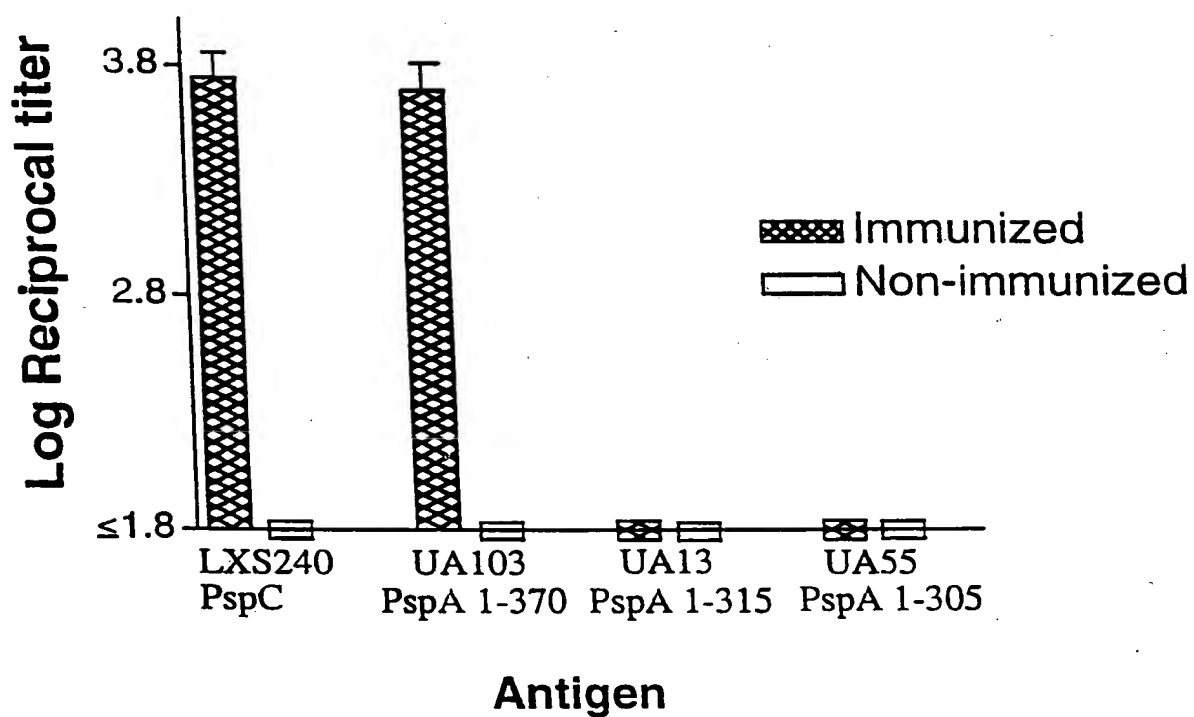
**FIG.6A**



**FIG.6B**



Western Immunoblot of pneumococcal lyates  
 Panel A was developed with anti-PspC polyclonal serum and panel B was developed with anti-PspA monoclonal antibody, Xi126. The molecular weight markers are indicated. Cross-reaction of the polyclonal serum to PspC is observed with all strains tested.



Level of antibody reactive to PspC and PspA fragments present in the sera of mice immunized with PspC. Each bar represents the mean of the log reciprocal titer and upperbound of standard error of sera from five mice. The limit of detection of the log reciprocal antibody titer is 1.8.

**FIG.7**

13/50

## FIG.8A

Other Formats:

Links:

LOCUS CAA05158 539 aa BCT 14-OCT-1997

DEFINITION SpsA protein.

ACCESSION CAA05158

PID g2576331

VERSION CAA05158.1 GI:2576331

DBSOURCE embl locus SPSPSA2, accession AJ002054.1

KEYWORDS

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (residues 1 to 539)

AUTHORS Hammerschmidt,S.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1997) Hammerschmidt S., Microbiology, Division  
Microbial Pathogenesis, National Research Centre for Biotechnology,  
Spielmannstrasse 7, 38106 Braunschweig, GERMANY

REFERENCE 2 (residues 1 to 539)

AUTHORS Hammerschmidt,S., Talay,S.R., Brandtzaeg,P. and Chhatwal,G.S.

TITLE SpsA, a novel pneumococcal surface protein with specific binding to  
secretory immunoglobulin A and secretory component

JOURNAL Mol. Microbiol. 25 (6), 1113-1124 (1997)

MEDLINE 98010350

FEATURES Location/Qualifiers

source 1..539

/organism="Streptococcus pneumoniae"

/strain="type 2"

/db\_xref="taxon:1313"

Protein 1..539

/function="IgA binding protein"

/product="SpsA protein"

CDS 1..539

/db\_xref="SPTREMBL:O33741"

/coded\_by="AJ002054.1:1..1620"

/transl\_table=11

ORIGIN

1 mfaskserkv hysirkfsig vasvavaslv mgsvvhaten egstqaatss nmaktehrka  
61 akqvvdyeie kmlreiqldr rkhtqnvaln iklsaiktky lrelnvleek skdelpseik  
121 akldaafekf kkdltkpgek vaeakkkvee akkkaedqke edrnyptnt ykteleiaie  
181 fdvkvkeael elvkeeakef megtikqak ekveskkaea trleniktdr kkaeeeaark  
241 aaedkvkek paeqpqpapa tqpekpapkp ekpaeqpkae ktddqqaeed yarrseeeyn  
301 rltqqppkt ekpaqpstpk tgwkqengmw yfyntdgsma tglwlnngsw yylnangama  
361 tglwlnngsw yylnangama tglwlnngsw yylnangama tglwlyngsw yylnsngama

14/50

***FIG. 8B***

421 tglqyngsw yynangdma tglqnnngsw yynangdma tglqyngsw yynangdma  
481 tgwvkdgdtw yyletsgamk asqwfksdk wyyvhgssal ainttvygyg vnangewvn

//

the above report in format

Other Formats:  
Links:

*FIG.8C*

LOCUS SPSPSA2 1620 bp DNA BCT 14-OCT-1997

DEFINITION Streptococcus pneumoniae SIgA binding.

ACCESSION AJ002054

NID g2576330

VERSION AJ002054.1 GI:2576330

KEYWORDS SpsA protein.

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 1620)

AUTHORS Hammerschmidt,S.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1997) Hammerschmidt S., Microbiology, Devision  
Microbial Pathogenesis, National Research Centre for Biotechnology,  
Spielmannstrasse 7, 38106 Braunschweig, GERMANY

REFERENCE 2 (bases 1 to 1620)

AUTHORS Hammerschmidt,S., Talay,S.R., Brandtzaeg,P. and Chhatwal,G.S.

TITLE SpsA, a novel pneumococcal surface protein with specific binding to  
secretory immunoglobulin A and secretory component

JOURNAL Mol. Microbiol. 25 (6), 1113-1124 (1997)

MEDLINE 98010350

FEATURES Location/Qualifiers

source 1..1620

/organism="Streptococcus pneumoniae"

/strain="type 2"

/db\_xref="taxon:1313"

CDS 1..1620

/function="IgA binding protein"

/codon\_start=1

/transl\_table=11

/product="SpsA protein"

/protein\_id="CAA05158.1"

/db\_xref="PID:e354783"

/db\_xref="PID:g2576331"

/db\_xref="GI:2576331"

/db\_xref="SPTREMBL:O33741"

/translation="MFASKSERKVHYSIRKFSIGVASVAVASLVMGSVVHATENEGST

QAATSSNMAKTEHRKAAKQVVDEYIEKMLREIQLDRRKHTQNVALNIKLSAIKT  
KYLR

16/50

FIG. 8D

ELNVLEEKSKDELPSEIKAKLDAAFEKFKKDTLKPGEKVAEAKKKVEEAKKKAE  
DQKE

EDRRNYPTNTYKTLELEIAEFDVKVKEAELELVKEEAKEFRNEGTIKQAKEKVES  
KKA

EATRLNIKTDRKKAEEEEAKRKAAEEDKVKEKPAEQPPAPATQPEKPAPKPEK  
PAEQ

PKAEKTTDDQQAEEYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGM  
WYFYN

TDGSMATGWLQNNGSWYYLNANGAMATGWLQNNGSWYYLNANGSMATGW  
LQNNGSWYY

LNANGAMATGWLQYNGSWYYLNSNGAMATGWLQYNGSWYYLNANGDMAT  
GWLQNNGSW

YYLNANGDMATGWLQYNGSWYYLNANGDMATGWVKDGDWYYLETSGAM  
KASQWFKVS

DKWYYVHGSSALAINTTVYGYGVNANGWVN"

repeat\_region 961..1564

/rpt\_type=DIRECT

BASE COUNT 625 a 294 c 370 g 331 t

ORIGIN

1 atgtttgcat caaaaagcga aagaaaagta cattattcaa ttcgtaaatt tagtattgga  
61 gtagctagtg tagctgttgc cagtcttgtt atgggaagtg tgggtcatgc gacagagaac  
121 gagggaaagta cccaagcagc cacttcttct aatatggcaa agacagaaca taggaaagct  
181 gctaaacaag tcgtcgatga atatatagaa aaaatgttga gggagattca actagataga  
241 agaaaacata cccaaaatgt cgccttaaac ataaagtga gcgcaattaa aacgaagtat  
301 ttgcgtgaat taaatgtttt agaagagaag tcgaaagatg agttgccgtc agaaataaaa  
361 gcaaagttag acgcagcitt tgagaagttt aaaaaagata cattgaaacc aggagaaaag  
421 gtagcagaag ctaagaagaa ggttgaagaa gctaagaaaa aagccgagga tcaaaaagaa  
481 gaagatcgtc gtaactaccr aaccaatact tacaaaacgc ttgaactga aattgctgag  
541 ttcgatgtga aagttaaaga agcggagctt gaactagtaa aagaggaagc taaagaattt  
601 cgaaacgagg gcacaattaa gcaagcaaaa gagaaagttg agagtaaaaa agctgaggct  
661 acaagggttag aaaacatcaa gacagatcgt aaaaaagcag aagaagaagc taaacgaaaa  
721 gcagcagaag aagataaagt taaagaaaaa ccagctgaac aaccacaacc agcgcgggct  
781 actcaaccag aaaaaccagc tccaaaacca gagaagccag ctgaacaacc aaaagcagaa  
841 aaaacagatg atcaacaagc tgaagaagac tatgctcgta gatcagaaga agaataaat  
901 cgcttgactc aacagcaacc gccaaaaact gaaaaaccag cacaaccatc tactccaaaa  
961 acaggctgga aacaagaaaa cggatgttgg tacttctaca atactgatgg ttcaatggca  
1021 acaggatggc tccaaaacaa cggttcatgg tactatctaa acgctaattg tgctatggcg  
1081 acaggatggc tccaaaacaa tggttcatgg tactatctaa acgctaattg ttcaatggca  
1141 acaggatggc tccaaaacaa tggttcatgg tactacctaa acgctaattg tgctatggcg



*FIG.8E*

1201 acaggatggc tccaatacaa tgggtcatgg tactacctaa acagcaatgg cgctatggcg  
1261 acaggatggc tccaatacaa tgggtcatgg tactacctca acgctaattg tgatatggcg  
1321 acaggatggc tccaaaacaa cgggtcatgg tactacctca acgctaattg tgatatggcg  
1381 acaggatggc tccaatacaa cgggtcatgg tattacctca acgctaattg tgatatggcg  
1441 acagggtggg tgaaagatgg agatacctgg tactatcttg aaacatcagg tgctatgaaa  
1501 gcaagccaat ggttcaaagt atcagataaa tgggtactatg tccatggctc aagtgcctt  
1561 gcaatcaaca caactgtata tgggtatgga gtcaatgccca atggtgaatg ggtaaacata

//

the above report in format

18/50

## FIG. 9A

Other Formats:

Links:

LOCUS CAA05159 581 aa BCT 14-OCT-1997

DEFINITION SpsA protein.

ACCESSION CAA05159

PID g2576333

VERSION CAA05159.1 GI:2576333

DBSOURCE embl locus SPSPSA47, accession AJ002055.1

KEYWORDS

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (residues 1 to 581).

AUTHORS Hammerschmidt,S.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1997) Hammerschmidt S., Microbiology, Division  
Microbial Pathogenesis, National Research Centre for Biotechnology,  
Spielmannstrasse 7, 38106 Braunschweig, GERMANY

REFERENCE 2 (residues 1 to 581)

AUTHORS Hammerschmidt,S., Talay,S.R., Brandtzaeg,P. and Chhatwal,G.S.

TITLE SpsA, a novel pneumococcal surface protein with specific binding to  
secretory immunoglobulin A and secretory component

JOURNAL Mol. Microbiol. 25 (6), 1113-1124 (1997)

MEDLINE 98010350

FEATURES Location/Qualifiers

source 1..581

/organism="Streptococcus pneumoniae"

/strain="type 47"

/db\_xref="taxon:1313"

Protein 1..581

/function="IgA binding protein"

/product="SpsA protein"

CDS 1..581


/db\_xref="SPTREMBL:O33742"

/coded\_by="AJ002055.1:1..1746"

/transl\_table=11

ORIGIN

1 mfaskserkv hysirkfsig vasvavaslv mgsvvhaten egstqaatss nmaktehrka  
61 akqvvdeyie kmlreiqldr rkhtqnvaln iklsaiktky lrelnvleek skdelpseik  
121 akldaafekf kkdtkpgek vaeakkkvee akkkaedqke edrrnyptnt yktleleiae  
181 fdvkvkeael elvkeeakes megtikqak ekveskkaea trleniktdr kkaeeekrk  
241 adaklikeanv atsdqgkpkg rakrgvpgel atpdkkenda kssdssvgee tlpsslksg  
301 kkvaeaekkv eaeakkakdq keedrnypt ntyktldlei aedsvkvkea elelvkeeak  
361 eprdeekikq akakveskka eatrlenikt drkkaeeek rkaaeedkvk ekpaeqpqpa



19/50

421 patqpekpap kpekpaepk aektdqqae edyarrsee ynriqqppp ktekpaqpft  
481 pktgwkqeng mwyfyntdgs matgwlyng swyylnangd matgwvkdgd twyyleasga  
541 mkasqwfkvs dkwyvngsg alavnttdg ygvnangewv n  
//

*FIG.9B*

the above report in format

20/50

## FIG.9C

Other Formats:

Links:

LOCUS SPSPSA47 1746 bp DNA BCT 14-OCT-1997

DEFINITION Streptococcus pneumoniae SIgA binding.

ACCESSION AJ002055

NID g2576332

VERSION AJ002055.1 GI:2576332

KEYWORDS SpsA protein.

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 1746)

AUTHORS Hammerschmidt,S.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1997) Hammerschmidt S., Microbiology, Division  
Microbial Pathogenesis, National Research Centre for Biotechnology,  
Spielmannstrasse 7, 38106 Braunschweig, GERMANY

REFERENCE 2 (bases 1 to 1746)

AUTHORS Hammerschmidt,S., Talay,S.R., Brandtzaeg,P. and Chhatwal,G.S.

TITLE SpsA, a novel pneumococcal surface protein with specific binding to  
secretory immunoglobulin A and secretory component

JOURNAL Mol. Microbiol. 25 (6), 1113-1124 (1997)

MEDLINE 98010350

FEATURES Location/Qualifiers

source 1..1746

/organism="Streptococcus pneumoniae"

/strain="type 47"

/db\_xref="taxon:1313"

CDS 1..1746

/function="IgA binding protein"

/codon\_start=1

/transl\_table=11

/product="SpsA protein"

/protein\_id="CAA05159.1"

/db\_xref="PID:e354769"

/db\_xref="PID:g2576333"

/db\_xref="GI:2576333"

/db\_xref="SPTREMBL:O33742"

/translation="MFASKSERKVHYSIRKFSIGVASVAVASLVMGSVVHATENEGST

QAATSSNMAKTEHRKAAKQVVDEYIEKMLREIQLDRRKHTQNVALNIKLSAIKT  
KYLR

21/50

*FIG.9D*

ELNVLEEKSKDELPSEIKAKLDAAFEKFKKDTLKPGEKVAEAKKKVEEAKKKAE  
DQKE

EDRRNYPTNTYKTLELEIAEFDVKVKEAELELVKEEAKESRNEGTIKQAKEKVES  
KKA

EATRLENIKTDRKKAEEEEAKRKADAKLKEANVATSDQGKPKGRAKRGVPGELA  
TPDKK

ENDAKSSDSSVGEETLPSSSLKSGKKVAEAEKKVEEAEKKAKDQKEEDRRNYPT  
NTYK

TLDLEIAESDVKVKEAELELVKEEAKEPRDEEKIKQAKAKVESKKAEATRLENIK  
TDR

KKAEKAAKRKAAEEDKVKEKPAEQPPAPATQPEKPAPKPEKPAEQPKAEKTDD  
QQA

EDYARRSEEEYNRLIQQQPPKTEKPAQPFPTKTGWKQENGWYFYNTDGSMAT  
GWLQY

NGSWYYLNANGDMATGWVKDGDWYYLEASGAMKASQWFKVSDKWYYVN  
GSGALAVNT

TVDGYGVNANGWVN"

repeat\_region 1447..1690

/rpt\_type=DIRECT

BASE COUNT 722 a 290 c 408 g 326 t

ORIGIN

1 atgtttgcat caaaaagcga aagaaaagta cattattcaa ttcgtaaatt tagtattgga  
61 gtagctagtg tagctgttgc cagtcttgtt atgggaagtg tggttcatgc gacagagaac  
121 gagggaaagta cccaagcagc cacttcttct aatatggcaa agacagaaca taggaaagct  
181 gctaaacaag tcgtcgatga atatatgaa aaaatgttga gggagattca actagataga  
241 agaaaacata cccaaaatgt cgccttaaac ataaagtga gcgcaattaa aacgaagtat  
301 ttgcgtgaat taaatgtttt agaagagaag tgcgaagatg agttgccgtc agaaataaaa  
361 gcaaagttag acgcagcttt tgagaagtt aaaaaagata cattgaaacc aggagaaaaag  
421 gtagcagaag ctaagaagaa ggttgaagaa gctaagaaaa aagccgagga tcaaaaagaa  
481 gaagatcgtc gtaactaccc aaccaatact taaaaacgc ttgaacttga aattgctgag  
541 ttcgatgtga aagttaaaga agcggagctt gaactagtaa aagaggaagc taaagaatct  
601 cgaaacgagg gcacaattaa gcaagcaaaa gagaaagttg agagtaaaaa agctgaggct  
661 acaaggttag aaaacatcaa gacagatcgt aaaaaagcag aagaagaagc taaacgaaaa  
721 gcagatgcta agttgaagga agctaagtga gcgacttcag atcaaggtaa accaaagggg  
781 cgggcaaaaac gaggagttcc tggagagcta gcaacacctg ataaaaaaga aaatgatgcg  
841 aagtcttcag attctagcgt aggtgaagaa acitctccaa gctcatccct gaaatcagga  
901 aaaaaggtag cagaagctga gaagaaggtt gaagaagctg agaaaaaagc caaggatcaa  
961 aaagaagaag atgccgtaa ttaccaacc aatacttaca aaacgcttga ccttgaaatt

1021 gctgagtcg atgtgaaagt taaagaagcg gagcttgaac tagtaaaaga ggaagctaag  
1081 gaacctcgag acgaggaaaa aattaagcaa gcaaaagcga aagttgagag taaaaaagct  
1141 gaggctacaa ggtagaaaa catcaagaca gatcgtaaaa aagcagaaga agaagctaaa  
1201 cgaaaagcag cagaagaaga taaagttaaa gaaaaaccag ctgaacaacc acaaccagcg  
1261 ccggctactc aaccagaaaa accagctcca aaaccagaga agccagctga acaaccaaaa  
1321 gcagaaaaaa cagatgatca acaagctgaa gaagactatg ctgtagatc agaagaagaa  
1381 tataatcgct tgattcaaca gcaaccgcca aaaactgaaa aaccagcaca accatttact  
1441 caaaaaacag gctggaaaca agaaaacggt atgtggtact tctacaatac tgatggttca  
1501 atggcaacag gatggctcca atacaacggt tcatggtatt acctcaacgc taatggtgat  
1561 atggcgacag gttgggtgaa agatggagat acctggtact atcttgaagc atcaggtgct  
1621 atgaaagcaa gccaatgggt caaagtatca gataaatggt actatgtcaa tggctcaggt  
1681 gcccttgacg tcaacacaac tgtagatggc tatggagtca atgccaatgg tgaatgggta  
1741 aactaa

//

*FIG. 9E*

the above report in format

23/50

## FIG.10A

ther Formats:  
Links:

LOCUS AAB70838 663 aa BCT 16-SEP-1997

DEFINITION choline binding protein A.

ACCESSION AAB70838

PID g2425109

VERSION AAB70838.1 GI:2425109

DBSOURCE locus AF019904 accession AF019904.1

### KEYWORDS

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Eubacteria; Firmicutes; Low G+C gram-positive bacteria;

Streptococcaceae; Streptococcus.

REFERENCE 1 (residues 1 to 663)

AUTHORS Rosenow,C., Ryan,P., Weiser,J.N., Johnson,S., Fontan,P.,  
Ortqvist,A. and Masure,H.R.

TITLE Contribution of novel choline-binding proteins to adherence,  
colonization and immunogenicity of Streptococcus pneumoniae

JOURNAL Mol. Microbiol. (1997) In press

REFERENCE 2 (residues 1 to 663)

AUTHORS Rosenow,C., Ryan,P., Weiser,J.N., Johnson,S., Fontan,P.,  
Ortqvist,A. and Masure,H.R.

TITLE Direct Submission

JOURNAL Submitted (18-AUG-1997) Infectious Diseases, St. Jude Children's  
Research Hospital, 322 North Lauderdale, Memphis, TN 38105, USA

COMMENT Method: conceptual translation supplied by author.

### FEATURES

Location/Qualifiers

source 1..663

/organism="Streptococcus pneumoniae"

/strain="R6x"

/db\_xref="taxon:1313"

Protein <1..663

/product="choline binding protein A"

/name="CbpA"

CDS 1..663

/gene="cbpA"

/coded\_by="AF019904.1:<1..1992"

/transl\_table=11

### ORIGIN

1 enegstqaat ssnmaktehr kaakqvvey iekmlreigl drkhtqnva lniklsaitk  
61 kylreinvle ekskdelpse ikakldaafe kfkddtlkpg ekvaeakkkv eeakkkaedq  
121 keedrmnypt ntyktlelei aefdvkvkea elelvkeeak esnegtikq akekveskka  
181 eatrlenikt drkkaeeek rkadakilkea nvatsdqqkp kgrakrgvpg elatpdkken  
241 dakssdsvg eetlpssslk sgkkvaeaek kveeaekkak dqkeedrmny ptntyktldl  
301 eiaesdvkvk eaelelvkee akeprdeeki kqakakvesk kaeatleni ktdrkkaeee

***FIG.10B***

361 akrkaeedk vkekpaepq papatqpekp apkpekpaeq pkaektdddq aeedyarrse  
421 eeynrltqq ppktekpaqp stpktgwkqe ngmwyfyntd gsmatgwlqn ngswyylnan  
481 gamatgwlqn ngswyylnan gsmatgwlqn ngswyylnan gamatgwlqy ngswyylnsn  
541 gamatgwlqy ngswyylnan gdmattgwlqn ngswyylnan gdmattgwlqy ngswyylnan  
601 gdmattgwkvd gdtwyyleas gamkasqwk vsdkwyvng sgalavnttv dgygvnange  
661 wvn

//

the above report in format



25/50

*FIG.10C*

Other Formats:

Links:

LOCUS AF019904 2480 bp DNA BCT 22-SEP-1997  
DEFINITION Streptococcus pneumoniae choline binding protein A (cbpA) gene,  
partial cds.

ACCESSION AF019904

NID g2425108

VERSION AF019904.1 GI:2425108

KEYWORDS

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Eubacteria; Firmicutes; Low G+C gram-positive bacteria;

Streptococcaceae; Streptococcus.

REFERENCE 1 (bases 1 to 2480)

AUTHORS Rosenow,C., Ryan,P., Weiser,J.N., Johnson,S., Fontan,P.,  
Ortqvist,A. and Masure,H.R.

TITLE Contribution of novel choline-binding proteins to adherence,  
colonization and immunogenicity of Streptococcus pneumoniae

JOURNAL Mol. Microbiol. (1997) In press

REFERENCE 2 (bases 1 to 2480)

AUTHORS Rosenow,C., Ryan,P., Weiser,J.N., Johnson,S., Fontan,P.,  
Ortqvist,A. and Masure,H.R.

TITLE Direct Submission

JOURNAL Submitted (18-AUG-1997) Infectious Diseases, St. Jude Children's  
Research Hospital, 322 North Lauderdale, Memphis, TN 38105, USA

FEATURES Location/Qualifiers

source 1..2480

/organism="Streptococcus pneumoniae"

/strain="R6x"

/db\_xref="taxon:1313"

gene <1..1992

/gene="cbpA"

CDS <1..1992

/gene="cbpA"

/note="CbP A"

/codon\_start=1

/transl\_table=11

/product="choline binding protein A"

/protein\_id="AAB70838.1"

/db\_xref="PID:g2425109"

/db\_xref="GI:2425109"

/translation="ENEGSTQAATSSNMAKTEHRKAAKQVVDEYIEKMLREIQLDRRK"

*FIG. 10D*

26/50

HTQNVALNIKLSAIKTKYLRELVLEEKSKDELPSSEIKAKLDAAFEKFKKDTLKP  
GEK

VAEAKKKVEEAKKKAEDQKEEDRRNYPTNTYKTLLELEIAEFDVKVKEAELELV  
KEEAK

ESRNEGTIKQAKEKVESKKAETRLENIKTDRKKAEEEEAKRKADAKLKEANVAT  
SDQG

KPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSSSLKSGKKVAEAEKKV  
EEAE

KKAKDQKEEDRRNYPTNTYKTLDLLEIAESDVKVKEAELELVKEEAKEPRDEEKI  
KQAK

AKVESKKAETRLENIKTDRKKAEEEEAKRKAAEEDKVKEKPAEQPPAPATQPE  
KPAP

KPEKPAEQPKAEKTDDQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTG  
WKQE

NGMWYFYNTDGSMATGWLQNNGSWYYLNANGAMATGWLQNNGSWYYLNA  
NGSMATGWL

QNNGSWYYLNANGAMATGWLQYNGSWYYLNSNGAMATGWLQYNGSWYYL  
NANGDMATG

WLQNNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWVKDGDWTWY  
YLEASGAMK

ASQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGWEVWV"

BASE COUNT 933 a 452 c 575 g 520 t

ORIGIN

1 gaaaacgaag gaagtaccca agcagccact tcttctaata tggcaaagac agaacatagg  
61 aaagctgcta aacaagtcgt cgatgaatat atagaaaaaa tgttgaggga gattcaacta  
121 gatagaagaa aacataccca aaatgtcgcc ttaaacataa agttgagcgc aattaaaacg  
181 aagtatttgc gtgaattaaa tgttttagaa gagaagtcga aagatgagtt gccgtcagau  
241 ataaaagcaa agttagacgc agcttttag aagtttaaaa aagatacatt gaaaccagga  
301 gaaaaggtag cagaagctaa gaagaagggt gaagaagcta agaaaaagc cgaggatcaa  
361 aaagaagaag atcgtcgtaa ctacccaacc aatacttaca aaacgcttga acttgaaat  
421 gctgagttcg atgtgaaagt taaagaagcg gagcttgaac tagtaaaaga ggaagctaaa  
481 gaatctcgaa acgagggcac aattaagcaa gcaaaagaga aagttgagag taaaaaagct  
541 gaggctacaa ggtagaaaa catcaagaca gatcgtaaaa aagcagaaga agaagctaaa  
601 cgaaaagcag atgctaagtt gaaggaagct aatgtagcga cttcagatca aggtaaacca  
661 aaggggcggg caaaacgagg agttcctgga gagctagcaa cacctgataa aaaagaaaat  
721 gatgcgaagt cttcagattc tagcgtaggt gaagaaactc ttccaagctc atccctgaaa

*FIG.10E*

781 tcaggaaaaa aggtagcaga agctgagaag aagggtgaag aagctgagaa aaaagccaag  
 841 gatcaaaaag aagaagatcg ccgtaactac ccaaccaata cttacaaaac gcttgacctt  
 901 gaaattgctg agtccgatgt gaaagttaaa gaagcggagc ttgaactagt aaaagaggaa  
 961 gctaaggaac ctcgagacga ggaaaaaatt aagcaagcaa aagcgaaagt tgagagtaaa  
 1021 aaagctgagg ctacaagggt agaaaacatc aagacagatc gtaaaaaagc agaagaagaa  
 1081 gctaaacgaa aagcagcaga agaagataaa gttaaagaaa aaccagctga acaaccacaa  
 1141 ccagcgccgg ctactcaacc agaaaaacca gtcctaaaac cagagaagcc agctgaacaa  
 1201 ccaaaagcag aaaaaacaga tgatcaacaa gctgaagaag actatgctcg tagatcagaa  
 1261 gaagaatata atcgcttgac tcaacagcaa ccgcaaaaaa ctgaaaaacc agcacaacca  
 1321 tctactcaa aaacaggctg gaaacaagaa aacgggtatgt ggtacttcta caatactgat  
 1381 ggttcaatgg caacaggatg gctccaaaac aacgggtcat ggtactatct aaacgcta  
 1441 ggtgctatgg cgacaggatg gctccaaaac aatgggtcat ggtactatct aaacgcta  
 1501 ggttcaatgg caacaggatg gctccaaaac aatgggtcat ggtactacct aaacgcta  
 1561 ggtgctatgg cgacaggatg gctccaatac aatgggtcat ggtactacct aaacagcaat  
 1621 ggcgctatgg cgacaggatg gctccaatac aatgggtcat ggtactacct caacgcta  
 1681 ggtgatatgg cgacaggatg gctccaaaac aacgggtcat ggtactacct caacgcta  
 1741 ggtgatatgg cgacaggatg gctccaatac aacgggtcat ggtattacct caacgcta  
 1801 ggtgatatgg cgacagggtg ggtgaaagat ggagatacct ggtactatct tgaagcatca  
 1861 ggtgctatga aagcaagcca atgggtcaaa gtatcagata aatgggtacta tgtcaatggc  
 1921 tcagggtccc ttgcagtcaa cacaactgta gatggctatg gagtcaatgc caatggtgaa  
 1981 tgggtaaact aaacctaata taactagtta atactgactt cctgtaagaa cttttaaag  
 2041 tattccctac aaataccata tccttcagt agataatata cccttgtagg aagtttagat  
 2101 taaaaaataa ctctgtaatc tctagccgga ttatagcgc tagagactac ggagttttt  
 2161 tgatgaggaa agaattggcgg cattcaagag actctttaag agagttacgg gttttaact  
 2221 attaaagcctt ctccaattgc aagaggcttc aatctctgct aggggtgctag ctgcgaaat  
 2281 ggctccacgg agtttggcag cgccagatgt tccacggaga tagtgaggag cgaggccgcg  
 2341 gaattcacga actgcgacgt ttctccttt gaggttaatc aatcgtttca agtggtcga  
 2401 ggcgatcttc atcttgctt caaagggtcaa atcaggtagg atttctctg ttcaaagtt  
 2461 tatgggtggc ctggttgaag

//

the above report in format

28/50

## FIG. 11A

Other Formats:

Links:

LOCUS AAD00184 929 aa BCT 07-OCT-1996  
DEFINITION surface protein C.  
ACCESSION AAD00184  
PID g4097980  
VERSION AAD00184.1 GI:4097980  
DBSOURCE locus SPU72655 accession U72655.1  
KEYWORDS  
SOURCE Streptococcus pneumoniae.  
ORGANISM Streptococcus pneumoniae  
Eubacteria; Firmicutes; Low G+C gram-positive bacteria;  
Streptococcaceae; Streptococcus.  
REFERENCE 1 (residues 1 to 929)  
AUTHORS Brooks-Walter,A., Tart,R.C., Briles,D.E. and Hollingshead,S.K.  
TITLE The *pspC* gene encodes a second pneumococcal surface protein  
homologous to the protection-eliciting PspA protein of  
Streptococcus pneumoniae  
JOURNAL Unpublished  
REFERENCE 2 (residues 1 to 929)  
AUTHORS Brooks-Walter,A., Tart,R.C., Briles,D.E. and Hollingshead,S.K.  
TITLE Direct Submission  
JOURNAL Submitted (26-SEP-1996) Microbiology, University of Alabama at  
Birmingham, 19th Street South, Box 10, Birmingham, AL 35294-2170,  
USA  
FEATURES Location/Qualifiers  
source 1..929  
/organism="Streptococcus pneumoniae"  
/strain="EF6796"  
/db\_xref="taxon:1313"  
Protein 1..929  
/product="surface protein C"  
CDS 1..929  
/gene="pspC"  
/coded\_by="U72655.1:319..3108"  
/transl\_table=11

### ORIGIN

1 mfaskserkv hysirkfsig vasvavaslf lggvvhaegv rsgnnltvts sgqdiskkya  
61 deveshlesi lkdvkknkk vqhtqnvqli tklseikkky lydlkvnvl eaelstktke  
121 tkekltatfe qfkcdtlpte pekkvaeaqk kveeakkkae dqkekdrmy ptityktlel  
181 eiaesdvevk kaelvlkvk akesqdeeki kqaeavesk qaeatrlkki ktdreeakrk  
241 adaklkeave knvatseqdk pkrrakrgvs gelatpdkke ndakssdssv geetlpspsl  
301 nmanesqteh rkdvdeyikk mlseiqlrr khtqnvlni klsaiktkyl yelsvikens  
361 kkeeltsktk aeltaafeqf kkdtkpekk vaeakkvee akkkakdqke edrnyptnt  
421 ykteleiae sdvkvkael elvkeeanes meekikqak ekveskkaea trlekiktdr

481 kkaeeekark aeeseekaae akqkvdaee yaleakiae yevqrlekel keidesdsed  
 541 yikeglrapl qskldtkkak lskleelsdk ideldaeiak levqlkdaeg nnnveayfke  
 601 glekttack aelekaeadl kkavdepetp apapqpapap ekpaekpapa pekpapapek  
 661 papapekpap apekpapape kpaptpetpk tgwkqengmw yfyntdgsma tgwlqnngsw  
 721 yylnsngama tgwlqnngsw yylnsngama tgwlqyngsw yylnangdma tgwlqyngsw  
 781 yylnangdma tgwfqyngsw yylnangdma tgwfqyngsw yylnangdma tgwlqyngsw  
 841 yylnsngamv tgwlqnngsw yylnangdma tdwvkdgdw yyleasgamk asqwfksdk  
 901 wyyvngsgal avnttvdsyr vnangewvn

//

the above report in format

*FIG.IIB*

30/50

Other Formats:

Links:

*FIG. 11C*

LOCUS SPU72655 3463 bp DNA BCT 02-JAN-1999  
DEFINITION Streptococcus pneumoniae surface protein C (pspC) gene, complete cds.

ACCESSION U72655

NID g4097979

VERSION U72655.1 GI:4097979

KEYWORDS

SOURCE Streptococcus pneumoniae.

ORGANISM Streptococcus pneumoniae

Eubacteria; Firmicutes; Low G+C gram-positive bacteria;  
Streptococcaceae; Streptococcus.

REFERENCE 1 (bases 1 to 3463)

AUTHORS Brooks-Walter,A., Tart,R.C., Briles,D.E. and Hollingshead,S.K.

TITLE The pspC gene encodes a second pneumococcal surface protein  
homologous to the protection-eliciting PspA protein of  
Streptococcus pneumoniae

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3463)

AUTHORS Brooks-Walter,A., Tart,R.C., Briles,D.E. and Hollingshead,S.K.

TITLE Direct Submission

JOURNAL Submitted (26-SEP-1996) Microbiology, University of Alabama at  
Birmingham, 19th Street South, Box 10, Birmingham, AL 35294-2170,  
USA

FEATURES Location/Qualifiers

source 1..3463

/organism="Streptococcus pneumoniae"

/strain="EF6796"

/db\_xref="taxon:1313"

gene 319..3108

/gene="pspC"

CDS 319..3108

/gene="pspC"

/codon\_start=1

/transl\_table=11

/product="surface protein C"

/protein\_id="AAD00184.1"

/db\_xref="PID:g4097980"

/db\_xref="GI:4097980"

/translation="MFASKSERKVHYSIRKFSIGVASVAVASLFLGGVVHAEGVRSGN

NLTVTSSGQDISKKYADEVESHLESILKDVKKNLKKVQHTQNVGLITKLSEIKKK  
YLY

31/50

*FIG.IID*

DLKVNVLSEAELTSKTKETKEKLTATFEQFKKDTLPTEPEKKVAEAQKKVEEAK  
KKA E

DQKEKDRRNYPTITYKTLLELEIAESDVEVKKA ELEVKKAKESQDEEKIKQAEA  
EVE

SKQAEATRLKKIKTDREEAKRKADAKLKEAVEKNVATSEQDKPKRRAKRGVSG  
ELATP

DKKENDAKSSDSSVGEETLPSPSLNMANESQTEHRKDVDEYIKKMLSEIQLDRR  
KHTQ

NVNLNIKLSAIKTKYLYELSVLKENSKEELTSKTKAELTAAFEQFKKDTLKPEK  
KVA

EAEKKVEEAKKKAKDQKEEDRRNYPTINTYKTLLELEIAESDVKVKEAELELVKEE  
ANES

RNEEKIKQAKEKVESKKAEATRLEKIKTDRKKAEEEEAKRKAEESEKKAAEAKQK  
VDAE

EYALEAKIAELEYEVQRLEKELKEIDESDSEDYLKEGLRAPLQSKLDTKKAKLSK  
LEE

LSDKIDELDAEIAKLEVQLKDAEGNNNVEAYFKEGLEKTTAEKKAELEKAEADL  
KKAV

DEPETPAPAPQPAPAPEKPAEKPAPEKPAPEKPAPEKPAPEKPAPEKPAPE  
KP

APTPETPKTGWKQENGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATG  
WLQNNGSW

YYLNSNGAMATGWLQYNGSWYYLNANGDMATGWLQYNGSWYYLNANGDM  
ATGWFQYNG

SWYYLNANGDMATGWFQYNGSWYYLNANGDMATGWLQYNGSWYYLNSNG  
AMVTGWLQN

NGSWYYLNANGSMATDWVKDGDWYYLEASGAMKASQWFKVSDKWYYVNG  
SGALAVNT

TVDSYRVNANGWVN"

BASE COUNT 1326 a 591 c 755 g 790 t 1 others  
ORIGIN

1 aagcttatgc ttgtcaataa tcacaaatat gtagatcata tcttgtag gacagtaaaa

61 catcctaatt actttttaa tattttacct gagttgattg gcttgacctt gttgagtc  
121 gcctatatga cttttgtttt agtttttcca gtttatgcag ttattttgta tgcacgaata  
181 gctgaagagg aaaagttatt acatgaagtt ataateccaa atggaagcat aaagagataa  
241 atacaaaatt cgatttatat acagttcata ttgaagtgat atagtaaggt taaagaaaaa  
301 atatagaagg aaataaacat gtttgcata aaaagcgaaa gaaaagtaca ttattcaatt  
361 cgtaaattta gtattggagt agctagtgta gctgttgcca gcttgttctt aggaggagta  
421 gtccatgcag aagggggttag aagtgggaat aacctcacgg ttacatctag tgggcaagat  
481 atatcgaaga agtatgctga tgaagtcgag tgcacatctag aaagtatatt gaaggatgct  
541 aaaaaaaatt tgaaaaaagt tcaacatacc caaatgtcg gcttaattac aaagtgagc  
601 gaaattaaaa agaagtattt gtatgactta aaagttaatg tttatcgga agctgagttg  
661 acgtcaaaaa caaaagaaac aaaagaaaag ttaaccgcaa ctttgagca gtttaaaaaa  
721 gatacattac caacagaacc agaaaaaaag gtagcagaag ctcaagaaga gggtgaagaa  
781 gctaagaaaa aagccgagga tcaaaaagaa aaagatcgcc gtaactacc aaccattact  
841 taaaaaacgc ttgaactga aattgctgag tccgatgtgg aagttaaaaa agcggagctt  
901 gaactagtaa aagtgaagc taaggaatct caagacgagg aaaaaattaa gcaagcagaa  
961 gcggaagtg agagtaaca agctgaggct acaaggtaa aaaaaatcaa gacagatcgt  
1021 gaagaagcta aacgaaaagc agatgctaag ttgaaggaaag ctgttgaaaa gaatgtagc  
1081 acttcagagc aagataaacc aaagaggcgg gcaaacgag gagtttctgg agagctagca  
1141 acacctgata aaaaagaaaa tgatgcgaag tcttcagatt ctgacgtagg tgaagaaact  
1201 ctccaagcc catcccttaa tatggcaaat gaaagtcaga cagaacatag gaaagatgc  
1261 gatgaatata taaaaaaaaa gttgagtgag atccaattag atagaagaaa acataccaa  
1321 aatgtcaact taaacataaa gttgagcgca attaaaacga agtatttga tgaattaagt  
1381 gttttaaag agaactcgaa aaaagaagag ttgacgtcaa aaaccaaac agagttaacc  
1441 gcagctttt agcagtttaa aaaagataca ttgaaaccag aaaaaaagg agcagaagct  
1501 gagaagaagg ttgaagaagc taagaaaaaa gccaggatc aaaaagaaga agatcgccgt  
1561 aactaccaa ccaatactta caaacgctt gaactgaaa ttgctgagtc cgatgtgaaa  
1621 gttaaagaag cggagcttga actagtaaaa gaggaagcta acgaatctc aaacgaggaa  
1681 aaaattaagc aagcaaaaga gaaagttgag agtaaaaaag ctgaggctac aaggttagaa  
1741 aaaatcaaga cagatcgtaa aaaagcagaa gaagaagcta aacgaaaagc agaagaatct  
1801 gagaaaaaag ctgctgaagc caaacaacaa gtggatgctg aagaatatgc tcttgaagct  
1861 aaaatcgctg agttggaata tgaagttcag agactagaaa aagagctcaa agagattgat  
1921 gagtctgact cagaagatta tcttaagaa ggcctccgtg ctctcttca atctaaattg  
1981 gatacaaaa aagctaaact atcaaaactt gaagagttga gtgataagat tgatgagta  
2041 gacgtgaaa ttgcaaaact tgaagttcaa cttaaagatg ctgaaggaaa caataatga  
2101 gaagcctact ttaaagaagg ttagagaaa actactgctg agaaaaaagc tgaattagaa  
2161 aaagctgaag ctgaccttaa gaaagcagtt gatgagccag aaactccagc tccggctcct  
2221 caaccagctc cagctccaga aaaaccagct gaaaaaccag ctccagctcc agaaaaacca  
2281 gctccagctc cagaaaaacc agctccagct ccagaaaaac cagctccagc tccagaaaaa  
2341 ccagctccag ctccagaaaa accagctcca actccagaaa ctccaaaaac aggctggaaa  
2401 caagaaaacg gtatgtgga ctctacaat actgatggtt caatggcaac aggctggctc  
2461 caaaacaatg gctcatgga ctacctcaac agcaatggcg ctatggcgac aggatggctc  
2521 caaaacaatg gctcatgga ctacctcaac agcaatggcg ctatggcgac aggatggctc  
2581 caatacaatg gttcatgga ctacctcaac gctaaggtg atatggcgac aggatggctc  
2641 caatacaatg gttcatgga ctacctcaac gctaaggtg atatggcgac aggatggctc  
2701 caatacaatg gttcatgga ctacctcaac gctaaggtg atatggcgac aggatggctc  
2761 caatacaatg gttcatgga ctacctcaac gctaaggtg atatggcgac aggatggctc



*FIG. IIF*

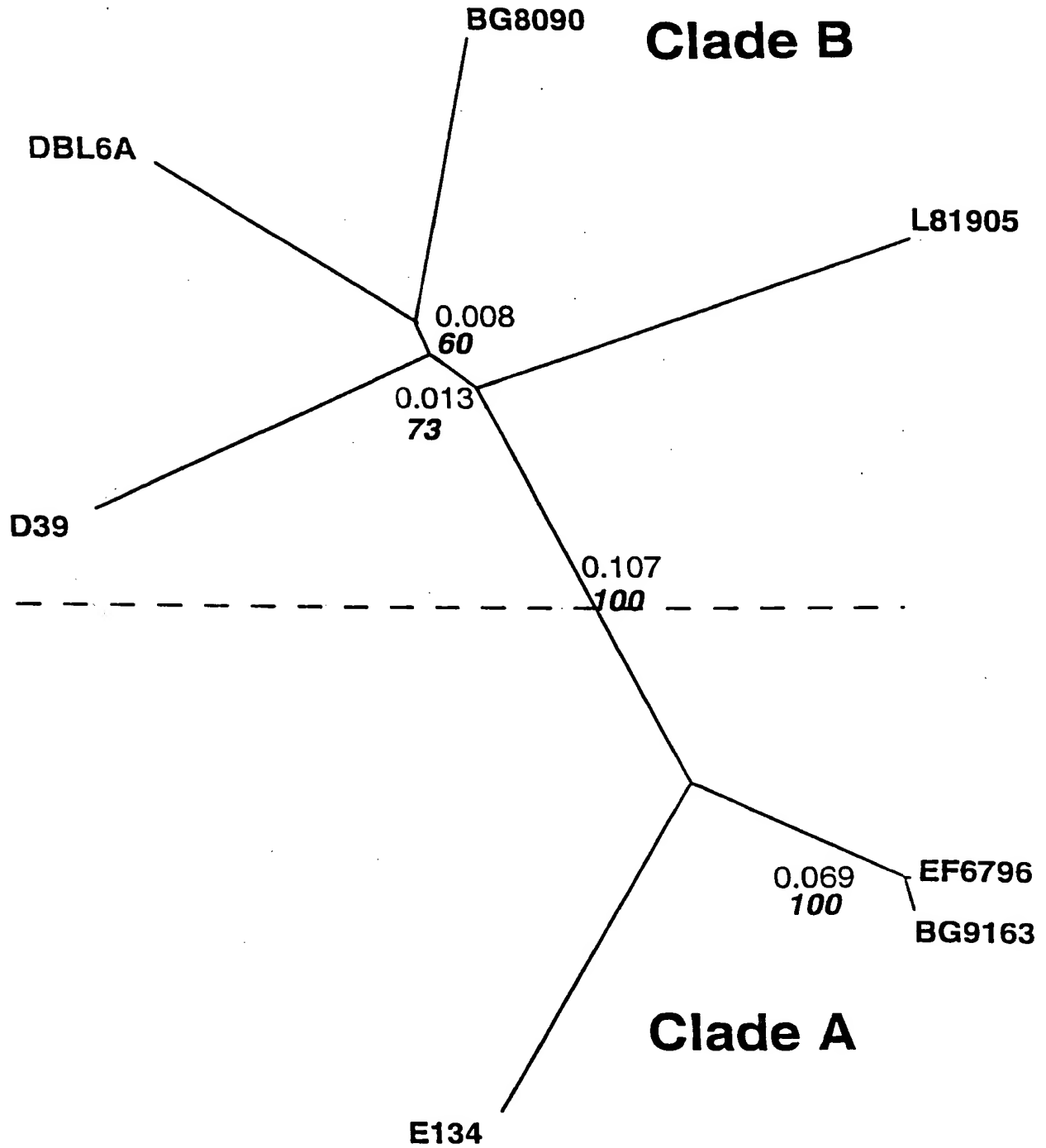
2821 caatacaatg gttcatggta ctacctaaac agcaatgggtg ctatggtaac aggatggctc  
2881 caaaacaatg gctcatggta ctacctaaac gctaacgggtt caatggcaac agattgggtg  
2941 aaagatggag atacctggta ctacttgaa gcatcagggtg ctatgaaagc aagccaatgg  
3001 ttcaaagtat cagataaatg gtactatgtc aatggctcag gtgcccttgc agtcaacaca  
3061 actgtagata gctatagagt caatgccaat ggtgaatggg taaactaaac ttaatataac  
3121 tagttaatac tgacttcctg taagaactct ttaaagtatt ccctacaaat accatatacct  
3181 tttagtagat aatataccct ttaggaagt ttagattaaa aaataactct gtaatctcta  
3241 gccggattta tagcgctaga gactacggag ttttttgat gaggaagaa tggcggcatt  
3301 caagagactc ttaagagag ttacgggttt taaactatta agctttctcc aattgcaaga  
3361 gggcttcaat ctctgctagg tgctagcttg cgaaatggct cccacggagt ttggcgcgc  
3421 cagatgttcc acggaggtag tgaggagcga ggccgcggaa ttc

//

the above report in format

34/50

*FIG. 12*



## UAB PspC SEQUENCES

EF6796	10	20	30	40	50	60	70
BG9763	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
E134	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
D39	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
DBL6A	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
BG8090	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
L81905	MFASKSERKVVHYS	IRKFSIGVASVA	SLFLGGVVHA	EGVR	SGNLTIT	SSGGQD	ISKYADEVES
EF6796	80	90	100	110	120	130	140
BG9763	-----HLESILKDV	KKNLKKVQ	-----HTQNVGL	IT	KLSEIT	KKKYL	YDL
E134	-----HLESILKDV	KKNLKKVQ	-----HTQNVGL	IT	KLSEIT	KKKYL	YDL
D39	-----DIKKMLSET	QEYIKKMLSE	-----IQD	KRKRHTQNVN	LN	IKLSA	IKT
DBL6A	-----AAKQVIV	DEYIKKMLSE	-----IQD	KRKRHTQNVN	LN	IKLSA	IKT
BG8090	-----KAAKQVIV	DEYIKKMLSE	-----IQD	KRKRHTQNVN	LN	IKLSA	IKT
L81905	KKLDSERDKARKEV	EEVKKIVGEISYAK	STKPKRHTIT	TVALLVN	ELNNIKN	TEYLNK	IVES
EF6796	150	160	170	180	190	200	210
BG9763	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
E134	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
D39	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
DBL6A	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
BG8090	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
L81905	ETKEKLTAT	FEQFKKDTL	-----PTEPEK	KVAE	AQKKV	EEA	KKKAEDQKE
EF6796	220	230	240	250	260	270	280
BG9763	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE
E134	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE
D39	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE
DBL6A	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE
BG8090	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE
L81905	NYPTITYKTLE	LEIAESDVEVK	KAEL	LELVKVKAKESQ	DEE	KIKQAE	AEVESKQAE

FIG. 13A

	290	300	310	320	330	340	350
EF6796	---	EEAKRKADAKLK	EEAVEKNVATSE	QDKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
BG9163	---	EEAKRKADAKLK	EEAVEKNVATSE	QDKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
E134	---	KAE	KAAEAKQVDTTKE	QGKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
D39	---	KAE	EESEK	QGKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
D131	---	KAE	EESEK	QGKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
DBL6A	ATRLNIKTDREKA	EEAKRKAAEEEV	---	QDKPKRRRAKRGVSG	ELAT	PDKKENDAKSS	DS
BG8090	TTRLNIKTDREKA	EEAKRKADAKE	---	QDESKRRVKG	QAT	LDKKENDAKSS	DS
81905	---	EAE	EEAKRRADAKE	QCKPKGRAKRGVSG	ELAT	PDKKENDAKSS	DS

	360	370	380	390	400	410	420
EF6796	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	DRRKHTIQNVN	NIKLSA	IKTKYLYELSVLKE	
BG9163	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	GRKHTPNVN	NIKLSA	IKTKYLYELSVLKE	
E134	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	RRKHTQNVN	NIKLSA	IKTKYLYELSVLKE	
D39	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	RRKHTQNVN	NIKLSA	IKTKYLYELSVLKE	
DBL6A	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	RRKHTQNVN	NIKLSA	IKTKYLYELSVLKE	
BG8090	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	RRKHTQNVN	NIKLSA	IKTKYLYELSVLKE	
B81905	SVGEETLPSP	SLNMANESQTEHRKDVDEYIKKMLSEIQ	LD	RRKHTQNVN	NIKLSA	IKTKYLYELSVLKE	

	430	440	450	460	470	480	490
EF6796	NSKKEELT	SKT	KAE	LTA	AFE	QF	KK
BG9163	NSKKEELT	SKT	KAE	LTA	AFE	QF	KK
E134	NSKKEELT	SKT	KAE	LTA	AFE	QF	KK
D39	-	-	-	-	-	-	-
DBL6A	-	-	-	-	-	-	-
BG8090	-	-	-	-	-	-	-
I81905	-	-	-	-	-	-	-

	500	510	520	530	540	550	560
EF6796	AESDVKVKEAELELVKEEAN	ESRNEEKIKQAK	EKVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA
BG9163	AESDVKVKEAELELVKEEAN	ESRNEEKIKQAK	EKVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA
E134	AESDVKVKEAELELVKEEAN	ESRNEEKIKQAK	EKVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA
D39	AESDVKVKEAELELVKEEAK	EPDNEEKIKQAK	EVESKKA	EATRL	ENIKT	DRKKA	EEEEAKRKA
DBL6A	AESDVKVKEAELELVKEEAK	ESRNEEKIKQAK	EVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA
BG8090	AESDVKVKEAELELVKEEAK	ESRNEEKIKQAK	EVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA
L81905	AESDVKVKEAELELVKEEAK	EPDNEEKIKQAK	EVESKKA	EATRL	EKIKT	DRKKA	EEEEAKRKA

**FIG. 13B**

EF6796  
 BG9163  
 E134  
 D39  
 DBL6A  
 BG8090  
 L81905

	570	580	590	600	610	620	630
EF6796	AEAKQKVDAAEY	AELEVEYVQRLE	KEKELKEI	DESDSE	DYKKEGLRAPLQ	SKLDT	KKAKLSKLEELS
BG9163	AEAKQKVDAAEY	AELEVEYVQRLE	KEKELKEI	DESDSE	DYKKEGLRAPLQ	SKLDT	KKAKLSKLEELS
E134	AEAKQKVDAAEY	AELEVEYVQRLE	KEKELKEI	DESDSE	DYKKEGLRAPLQ	SKLDT	KKAKLSKLEELS
D39	---	---	---	---	---	---	---
DBL6A	---	---	---	---	---	---	---
BG8090	---	---	---	---	---	---	---
L81905	---	---	---	---	---	---	---

EF6796  
 BG9163  
 E134  
 D39  
 DBL6A  
 BG8090  
 L81905

	640	650	660	670	680	690	700
EF6796	DKIDELDAEIAK	LEVQLKDAE	GNNNVEAY	FKEGLEKTT	AEKKAELEKAE	ADLKKAVDE	PEITPA
BG9163	DKIDELDAEIAK	LEVQLKDAE	GNNNVEAY	FKEGLEKTT	AEKKAELEKAE	ADLKKAVDE	PEITPA
E134	DKIDELDAEIAK	LEVQLKDAE	GNNNVEAY	FKEGLEKTT	AEKKAELEKAE	ADLKKAVDE	PEITPA
D39	---	---	---	---	---	---	---
DBL6A	---	---	---	---	---	---	---
BG8090	---	---	---	---	---	---	---
L81905	---	---	---	---	---	---	---

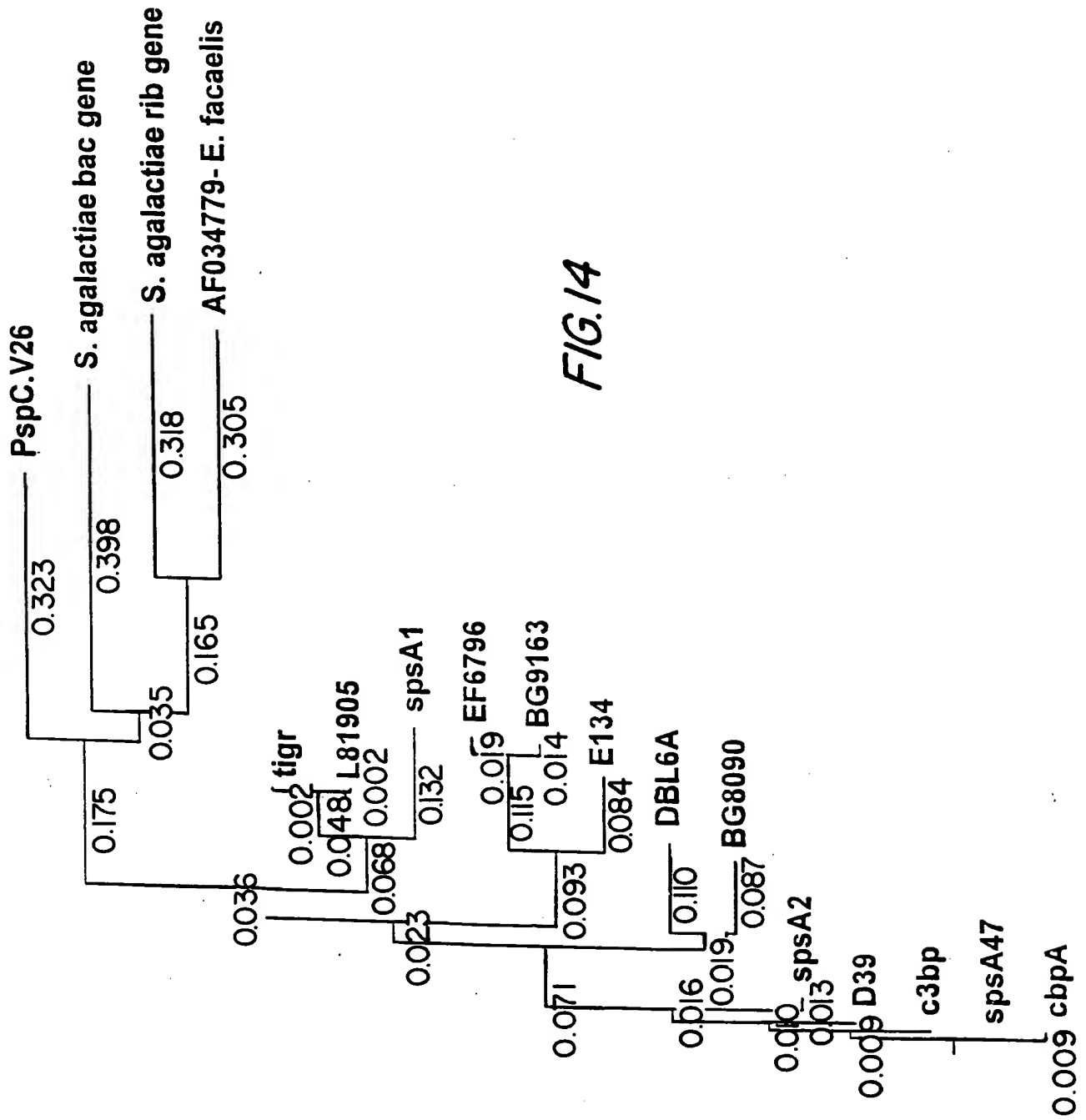
EF6796  
 BG9163  
 E134  
 D39  
 DBL6A  
 BG8090  
 L81905

	710	720	730	740	750	760	770
EF6796	APEKP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP
BG9163	APEKP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP
E134	APEKP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP	AEKPAPAP
D39	---	---	---	---	---	---	---
DBL6A	---	---	---	---	---	---	---
BG8090	---	---	---	---	---	---	---
L81905	---	---	---	---	---	---	---

EF6796  
 BG9163  
 E134  
 D39  
 DBL6A  
 BG8090  
 L81905

	780	790	800	810	820	830	840
EF6796	PEKPAPAPK	PETPET					
BG9163	PEKPAPAPK	PETPET					
E134	PEKPAPAPK	PETPET					
D39	---	---	---	---	---	---	---
DBL6A	---	---	---	---	---	---	---
BG8090	---	---	---	---	---	---	---
L81905	---	---	---	---	---	---	---

FIG.13C





39/50

## FIG.15A

LOCUS PSPC.V26 730 AA PROT SYN  
DEFINITION Streptococcus pneumoniae PspC protein from strain V26  
ACCESSION -  
KEYWORDS -  
SOURCE -  
FEATURES Location/Qualifiers  
PROPEP <1..731  
/note="3 to 2349 of V26.pspC (translated)"

ORIGIN -  
1 MFASKSERKV HYSIRKFSIG VASVVVASLF LGGVVHAEV RRGNNLTVTS SGDEVESHYQ  
61 SILEKVRKSL EKDRHTQNVD LIKKLQDIKR TYLYNLKEKP EAELTSKTNK ELDAAFEKFK  
121 KEPELTKKLA EAEKKAKDQK EEDHRNYPTN TYKTIELEIA EAEVGVAKAE LELAQAQVQI  
181 PQDTEKINAA KAKVEAAKSN VKKLEKIKSD IEKTYLYKLD NSTKETPKSR VRRNSPQVGD  
241 SRELKETIDK AKETLSTYMV TRLTKLDPSV FWFADLLMDA KKVVEEYKTK LEDASDKKSV  
301 EDLRKEAEGK IESLIVTHQN REKENQPAPQ PGGQAGGSMV VPPVTQTTPS TSQSPGQKAT  
361 EAEKKKLQDL IRQFQEALNK LDDETKTVPD GAKLTGEAGK AYNETRTRYAK EVVDKSKKLL  
421 SQTAVTMDEL AMQLTKLND MSKLKEAKAK LVPEVKPQPE NPEPKPQPEG EKPSVPDINQ  
481 EKEKAKLAIA TYMSKILDDI KKHHLKKEKH HQIVALIKDL DKLRKQALSE IDNVNTKVEI  
541 ENTVHKVFAD MDTVVTKFQK GLIQNTPQVP EAQRAQRYQR FQIHQKAPDT PQVPEAPKSP  
601 EVPKVPEAPK APDTPQVPEA PKSPEVPKVS DTPKAPDTPQ VPEAPKSPEV PKVPEAPKAP  
661 DTPQVPEAPK SPEVPKVPDT PKAPDTPQVP EAPKAPDTPQ IPEAPAPETP APAPEAPKTG  
721 WKQENGMWKG

//

LOCUS V26DNA.PSP 2349 BP SS-DNA SYN  
DEFINITION Streptococcus pneumoniae PspC gene from strain V26  
ACCESSION -  
KEYWORDS -  
SOURCE -  
BASE COUNT 937 A 457 C 505 G 450 T 0 OTHER  
ORIGIN -

## FIG.15B

1 AATTGGCCCT TCGACGAATA GCTGAAGAGG AAAAGCTATT ACAIGAAGTT ATAATCCCAA  
61 ATGGAAGCAT AAAGAGATAA ATACAAAATT CGATTATAT ACAGTTCATA TTGAAGTGAT  
121 ATAGTAAGGC TAAAGAAAAA ATATAGAAGG AAATAACAT GTTGGCATCA AAAAGCGAAA  
181 GAAAAGTACA TTATTCAAT CGTAAATTTA GTATTGGAGT AGCTAGTGTA GTAGTTGCTA  
241 GTTGTCTT AGGAGGAGTA GTTCACCGCAG AAGAGGTTAG AAGAGGGAAT AACCTCACGG  
301 TTACATCTAG TGGGGATGAA GTCGAGTCGC ATTATCAAAG TATATTGGAG AAGGTCAGAA  
361 AAAGTTTGGG AAAAGATCGA CATACCCAA ATGTCGACTT AATCAAAAAG TTGCAAGACA  
421 TTAAGAGAAC GTATTGTAT AATTAAAAAG AGAAGCCGGA AGCTGAGTTG ACGTCAAAAA  
481 CAAATAAAGA GTTAGACGCA GCTTTTGAGA AGTTAAAAA AGAACCAGAA CTTACTAAAA  
541 AATTAGCAGA AGCTGAGAAA AAGCCCAAGG ATCAAAAAGA AGAAGATCAC CGTAACCTACC  
601 CAACCAATAC TTACAAAACA ATCGAACTGG AATTGCGGA AGCAGAAGTA GGGGTCGCCA  
661 AGGCAGAGCT TGAGCTTGCA CAAGCTCAAG TCCAAATACC TCAAGATACT GAGAAAATTA  
721 ATGCTGCTAA AGCTAAAGTA GAAGCTGCTA AAGTAATGT TAAAAAATA GAAAAAATTA  
781 AATCAGATAT TGAaaaaaCG TATTGTATA AATTAGATAA CTCaaccCAA GAAACGCCAA  
841 AATCTAGAGT GCGAAGAAAT TCTCCGCAAG TAGGCGATTG GAGAGAACTT AAGGAAACGA  
901 TAGACAAAGC GAAAGAACT CTGCTACCT ATATGGTAAC TCGTTTAACG AAGCTGGATC  
961 CATCTGTTT TTGGTTTGA GATCTCTTA TGGATGCTAA GAAAGTTGTG GAAGAATACA  
1021 AGACAAAAT AGAGGATGCT TCAGATAAAA AATCGGTAGA AGACTTGCGA AAGGAAGCAG  
1081 AAGGAAAAAT AGAGTCTCTT ATCGTGACTC ACCAAAATAG AGAAAAAGAA AACCAACCCAG  
1141 CACCCCAACC AGGAGGACAA GCAGGTGGT CAATGGTTGT ACCACGGTG ACGCAACACAC  
1201 CTCCATCAAC TTCCCAAAGT CCAGGACAAA AGGCGACCGA AGCTGAAAAAG AAAAAGTTAG  
1261 AAGACTTGAT TCGTCAATC CAAGAAGCCT TGAACAAACT AGACGATGAA ACAAAGACTG  
1321 TTCCAGATGG GGCTAAACTC ACAGGAGAAG CTGGAAGAAG CTATAATGAG ACTAGAACTT  
1381 ATGCGAAAGA AGTTGTTGAC AAGAGCAAGA AGCTTCTATC ACAGACAGCA GTGACAATGG  
1441 ATGAATTGGC AATGCAATTA ACCAAATTGA ACGATGCCAT GTCTAAATTG AAAGAAGCTA  
1501 AAGCGAAATT GGTACCAGAG GTTAAACCAC AGCCGGAAAA CCCAGAGCCA AAACCCACAAC  
1561 CAGAGGGTGA GAAACCAAGC GTACCAGATA TTAATCAGGA GAAAGAAAAA GCTAAACTTG  
1621 CTATAGCAAC ATACATGAGC AAGATTTTAG ATGATATAA GAAACATCAT CTGAAGAAAAG  
1681 AAAACATCA TCAGATTGTT GCTCTTATTA AGGACCTTGA TAACTTAGA AAGCAAGCAC  
1741 TTTCTGAAAT TGATAATGTA AATACCAAAG TAGAAATTGA GAATACAGTC CACAAGGTAT  
1801 TTGCAGACAT GGATACGGTT GTTACTAAAT TCCAAAAAGG CTTAATTCAG AACACACCGC  
1861 AGGTTCCAGA AGCCCAAAGA GCCCAGAGGT ACCAAAGGTT TCAGATACAC CAAAAGGCTC



41/50

1921 CGGACACACC GCAGGTTCCA GAAGCACCAA AGAGCCCAGA GTTACCAAAG GTTCCAGAAG  
1981 CACCAAAGGC TCCGGACACA CCGCAAGTTC CGGAAGCACC AAAGAGCCCA GAGGTACCAA  
2041 AGGTTTCAGA TACACCAAAG GCTCCGGACA CACCGCAGGT TCCAGAAGCA CCAAAGAGCC  
2101 CAGAGGTACC AAGGTTCCA GAAGCACCAA AGGCTCCGGA CACACCGCAA GTTCCGGAAG  
2161 CACCAAAGAG CCCAGAGGTA CCAAGGTTT CAGATACACC AAAGGCTCCG GACACACCGC  
2221 AGGTTCCAGA AGCACCAAAG GCTCCAGACA CACCGCAAAT TCCGGAAGCA CCAGCTCCAG  
2281 AAATCCGGC TCCAGCTCCA GAAGCTCCA AAACAGGCTG GAAACAAGAA AACGGTATGT  
2341 GGAAGGGCG

//

FIG.15C

42/50

FIG. 16A

LOCUS E134.DNA 2405 BP SS-DNA SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

CDS 187..2405

BASE COUNT 1046 A 386 C 505 G 468 T 0 OTHER

ORIGIN -

1 CGGCCGCCAG TGTGATGGAT ATCTGCAGAA TTCGCCCTTC GACGAATAGC TGAAGAGGAA  
61 AAGCTATTAC ATGAAGTTAT AATCCCAAAT GGAAGCATAA AGAGATAAAT ACAAATTCG  
121 ATTTATATAC AGTTCATATT GAAGTGATAT AGTAAGGTTA AAGAAAAAAT ATAGAAGGAA  
181 ATAAACATGT TTGCATCAAA AAGCGAAAGA AAAGTACATT ATTCAATTCG TAAATTTAGT  
241 ATTGGAGTAG CTAGTGTAGT AGTTGCTAGT CTTGTTATGG GAAGTGTGGT TCATGCGACG  
301 GAGAATGAGG GAATTACCCA AGTAGCCACT TCTTATAATA AGGCAAATGA AAGTCAGACA  
361 GAACATAGGA AAGCTGCTAA ACAAGTCGAT GAAGATATAA AAAAAATGTT GAGTGAGATC  
421 CAAGAATATA TAAAAAAAAT GTTGAGTGAG ATCCAATTAG ATAAAAGAAA ACATACCCAA  
481 AATGTCAACT TAAACAGAAA GTTGAGCGCA ATTCAAACGA AGTATTTGTA TGAATTAAGA  
541 GTTTTAAAAG AGAAGTCGAA AAAAGAAGAG TTGACGTCAA AAACAAAAAA AGAGTTAGAC  
601 GCAGCTTTTG AGAAGTTTAA AAAAGAACCA GAACTTACTA AAAAAATTAGC AGAAGCTAAA  
661 CAAAAAGCCA AGGCTCAAAA AGAAGAAGAT TTCCGTAACCT ACCCAACCAA TACTTACAAA  
721 ACGCTTGAAC TTGAAATTGC TGAGTTTCGAT GTGAAAGTTA AAGAAGCGGA GCTTGAACCTA  
781 GTAAAAGAGG AAGCTAAACC CCGAAACGAG GAAAAAATTA AGCAAGCAAA AGCGAAAGTT  
841 GAGAGTAAAA AAGCTGAGGC TACAAGGTTA GAAGAAATCA AGACAGAACG TAAAAAGCA  
901 GAAGAAGAAG CTAAACGAAA AGCAGAAGAA TCTGAGAAAA AAGCTGCTGA AGCCAAACAA  
961 AAAGTGGATA CTAAAGAGCA AGGTAAACCA AAGAGGCGGG CAAAACGAGG  
AGTTTCTGGA  
1021 GAGCTAGCAA CACCTGATAA AAAAGAAAAT GATGCGAAGT CTTGAGATTC TAGCGTAGGT  
1081 GAAGAACTC TTCCAAGCCC ATCCCTTAAT ATGGCAAATG AAAGTCAGAC AGAACATAGG  
1141 AAAGATGTCG ATGAATATAT AAAAAAATG TTGAGTGAGA TCCAATTAGA TAGAAGAAAA  
1201 CATACCCAAA ATGTCAACTT AACATAAAG TTGAGCGCAA TTAAAACGAA GTATTTGTAT  
1261 GAATTAAGTG TTTTAAAAGA GAACTCGAAA AAAGAAGAGT TGACGTCAA AACCAGGCA  
1321 GAGTTAACCG CAGCTTTTGA GCAGTTTAAA AAAGATACAT TGAACCAGA AAAAAAGGTA  
1381 GCAGAAGCTG AGAAGAAGGT TGAAGAAGCT AAGAAAAAAG CCAAGGATCA  
AAAAGAAGAA  
1441 GATCGCCGTA ACTACCCAAC CAATACTTAC AAAACGCTTG AACTTGAAAT TGCTGAGTCC  
1501 GATGTGAAAG TTA AAAAAGC GGAGCTTGAA CTAGTAAAAG AGGAAGCTAA CGAATCTCGA  
1561 AACGAGGAAA AAATTAAGCA AGCAAAAGAG AAAGTTGAGA GTAAAAAGC TGAGGCTACA  
1621 AGGTTAGAAA AAATCAAGAC AGATCGTAAA AAAGCAGAAG AAGAAGCTAA ACGAAAAGCA  
1681 GAAGAATCTG AGAAAAAAGC TCTGAAGCC AAACAAAAAG TGGATGCTGA AGAATATGCT  
1741 CTTGAAGCTA AAATCGCTGA GTTGGAATAT GAAGTTTACA GACTAGAAAA AGAGCTCAAA  
1801 GAGATTGATG AGTCTGACTC AGAAGATTAT CTTAAAGAAG GCCTCCGTGC TCCTCTTCAA  
1861 TCTAAATTGG ATACCAAAAA AGCTAACTA TCAAACTTG AAGAGTTGAG TGATAAGATT  
1921 GATGAGTTAG ACGCTGAAAT TGCAAACTT GAAGTTCAAC TTAAGATGC TGAAGGAAAC  
1981 AATAATGTAG AAGCCTACTT TAAAGAAGGT TTAGAGAAAA CTACTGCTGA GAAAAAGCT  
2041 GAATTAGAAA AAGCTGAAGC TGACCTTAAG AAAGCAGTTG ATGAGCCAGA AACTCCAGCT  
2101 CCGGCTCCTC AACCAGCTCC AGCTCCAGAA AAACCAGCTG AAAAACCAGC  
TCCAGCTCCA  
2161 GCTCCAGAAA AACCAGCTCC AGCTCCAGAA AAACCAGCTG AAAAACCAGC  
TGAAAAACCA  
2221 GCTGAAGAAC CAGCTGAAAA ACCAGCTCCA GCTCCAGAAA AACCAGCTCC  
AACTCCAGAA  
2281 AAACCAGCTC CAACTCCAGA AACTCCAAAA ACAGGCTGGA AACAAGAAAA  
CGGTATGTGG



43/50

2341 TACTTCTACA ATACTGATGG TTCAATGGCA ACAGGCTGGC.TCCAAAACAA TGGTTTCATGG  
2401 TACTA

//

*FIG.16B*

44/50

LOCUS D39.DNA 2774 BP SS-DNA SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

CDS 362..2467

BASE COUNT 1057 A 476 C 615 G 624 T 2 OTHER

ORIGIN -

*FIG.17A*

1 CCAAGCTATT AGGTGACACT ATAGAATACT CAAGCTATGC ATCAAGCTTA TGCTTGTCAA  
61 TAATCACAAA TATGTAGATC ATATCTTGTT TAGGACAGTA AAACATCCTA ATTACTTTTT  
121 AAATATTCTT CTGAGTTGA TTGGCTTGAC CTTGTTGAGT CATGCTTATG TGACTTTTGT  
181 TTTAGTTTTT CCAGTTTATG CAGTTATTTT GTATCGACGA ATAGCTGAAG AGGAAAAGCT  
241 ATTACATGAA GTTATAATCC CAAATGGAAG CATAAAGAGA TAAATACAAA ATTCGATTTA  
301 TATACAGTTC ATATTGAAGT AATATAGTAA GGTTAAAGAA AAAATATAGA AGGAAATAAA  
361 CATGTTTGCA TCAAAAAGCG AAAGAAAAGT ACATTATTCA ATTCGTAAAT TTAGTATTGG  
421 AGTAGCTAGT GTAGCTGTTG CCAGTCTTGT TATGGGAAGT GTGGTTCATG CGACAGAGAA  
481 CGAGGGAAGT ACCCAAGCAG CCACCTTCTT TAATATGGCA AAGACAGAAC ATAGGAAAGC  
541 TGCTAAACAA GTCGTCGATG AATATATAGA AAAAATGTTG AGGGAGATTC AACTAGATAG  
601 AAGAAAACAT ACCCAAAATG TCGCCTTAAA CATAAAGTTG AGCGCAATTA AAACGAAGTA  
661 TTTGCGTGAA TTAAATGTTT TAGAAGAGAA GTCGAAAGAT GAGTTGCCGT CAGAAATAAA  
721 AGCAAAGTTA GACGCAGCTT TTGAGAAGTT TAAAAAGAT ACATTGAAAC CAGGAGAAAA  
781 GGTAGCAGAA GCTAAGAAGA AGGTTGAAGA AGCTAAGAAA AAAGCCGAGG ATCAAAAAGA  
841 AGAAGATCGT CGTAACTACC CAACCAATAC TTACAAAACG CTTGAACTTG AAATTGCTGA  
901 GTTCGATGTG AAAGTTAAAG AAGCGGAGCT TGAAGTAGTA AAAGAGGAAG CTAAAGAATC  
961 TCGAAACGAG GGCACAATTA AGCAAGCAAA AGAGAAAGTT GAGAGTAAAA AAGCTGAGGC  
1021 TACAAGGTTA GAAAACATCA AGACAGATCG TAAAAAGCA GAAGAAGAAG CTAAACGAAA  
1081 AGCAGATGGT AAGTTGAAGG AAGCTAATGT AGCGACTTCA GATCAAGGTA  
AACCAAAGGG  
1141 GCGGGCAAAA CGAGGAGTTC CTGGAGAGCT AGCAACACCT GATAAAAAAG  
AAAATGATGC  
1201 GAAGTCTTCA GATTCTAGCG TAGGTGAAGA AACTCTTCCA AGCTCATCCC TGAAATCAGG  
1261 AAAAAAGGTA GCAGAAGCTG AGAAGAAGGT TGAAGAAGCT GAGAAAAAAG  
CCAAGGATCA  
1321 AAAAGAAGAA GATCGCCGTA ACTACCCAAC CAATACTTAC AAAACGCTTG ACCTTGAAAT  
1381 TGCTGAGTCC GATGTGAAAG TTAAAGAAGC GGAGCTTGAA CTAGTAAAAG  
AGGAAGCTAA  
1441 GGAACCTCGA GACGAGGAAA AAATTAAGCA AGCAAAAGCG AAAGTTGAGA  
GTAAAAAGC  
1501 TGAGGCTACA AGGTTAGAAA ACATCAAGAC AGATCGTAAA AAAGCAGAAG AAGAAGCTAA  
1561 ACGAAAAGCA GCAGAAGAAG ATAAAGTTAA AGAAAAACCA GCTGAACAAC CACAACCAGC  
1621 GCCGGCTACT CAACCAGAAA AACCAGCTCC AAAACCAGAG AAGCCAGCTG  
AACAACCAAA  
1681 AGCAGAAAAA ACAGATGATC AACAAGCTGA AGAAGACTAT GCTCGTAGAT CAGAAGAAGA  
1741 ATATAATCGC TTGACTCAAC AGCAACCGCC AAAAAGTAA AAACCAGCAC AACCATCTAC  
1801 TCCAAAAACA GGCTGGAAC AAGAAAACGG TATGTGGTAC TTCTACAATA CTGATGGTTC  
1861 AATGGCAACA GGATGGCTCC AAAACAACGG TTCATGGTAC TATCTAAACG CTAATGGTGC  
1921 TATGGCGACA GGATGGCTCC AAAACAATGG TTCATGGTAC TATCTAAACG CTAATGGTTC  
1981 AATGGCAACA GGATGGCTCC AAAACAATGG TTCATGGTAC TACCTAAACG CTAATGGTGC  
2041 TATGGCGACA GGATGGCTCC AATACAATGG TTCATGGTAC TACCTAAACA  
GCAATGGCGC  
2101 TATGGCGACA GGATGGCTCC AATACAATGG CTCATGGTAC TACCTCAACG  
CTAATGGTGA  
2161 TATGGCGACA GGATGGCTCC AAAACAACGG TTCATGGTAC TACCTCAACG

*FIG.17B*

CTAATGGTGA

2221 TATGGCGACA GGATGGCTCC AATACAACGG TTCATGGTAT TACCTCAACG CTAATGGTGA

2281 TATGGCGACA GGTTGGGTGA AAGATGGANA TACCTGGTAC TATCTTAAAG  
CATCAGGTGC

2341 TATGAAAGCA AGCCAATGGT TCAAAGTATC AGATAAATGG TACTATGTCA ATGGCTCAGG

2401 TGCCCTTGCA GTCAACACAA CTGTAGATGG CTATGGAGTC AATGCCAATG  
GTGAATGGGT

2461 AAATAAACC TAATATAACT AGTTAATACT GACTTCCTGT AAGAACTTTT TAAAGTATTC

2521 CCTACAAATA CCATATCCTT TCAGTAGATA ATATACCCTT GTAGGAAGTT TAGATTAAAA

2581 AATAACTCTG TAATCTCTAG CCGGATTTAT AGCGCTAGAG ACTACGGAGT TTTTTTGATG

2641 AGGAAAGAAT GGCGGCATTC AAGAGACTCT TTAAGAGAGT TACGGGTTTT AAATAATTAA

2701 GCCTTCTCCA ATTGCAAGAG GGCTTCAATC TCTGCTAGGG TGCTAGCTTG  
CGAAATGGCT

2761 CCACGGAGTT TNGC

//

46/50

LOCUS BG9163.DNA 2489 BP SS-DNA

SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

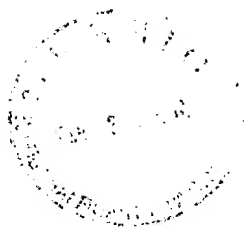
CDS 221..2489

BASE COUNT 1026 A 429 C 541 G 493 T 0 OTHER

ORIGIN -

FIG.18A

1 GATTGTATAC GACCACTATA GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT  
61 GGCCGCGGGT ATTCGACGAA TAGCTGAAGA GGAAAAGCTA TTACATGAAG TTATAATCCC  
121 AAATGGAAGC ATAAAGAGAT AAATACAACA TTCGATTTAT ATACAGTTCC TATTGAAGTG  
181 ATATAATAAG GTTAAAGAAA AAATATAGAA GGAAATAAAC ATGTTTGCAT CAAAAAGCGA  
241 AAGAAAAGTA CATTATTCAA TTCGTAAATT TAGTATTGGA GTAGCTAGTG TAGCTGTTGC  
301 CAGCTTGTTT TTAGGAGGAG TAGTCCATGC AGAAGGGGTT AGAAGTGGGA ATAACCTCAC  
361 GGTTACATCT AGTGGGCAAG ATATATCGAA GAAGTATGCT GATGAAGTCG AGTCGCATCT  
421 AGAAAGTATA TTGAAGGATG TCAAAAAAAA TTTGAAAAA GTTCAACATA CCCAAATGT  
481 CGGCTTAATT ACAAAGTTGA GCGAAATTAA AAAGAAGTAT TTGTATGACT TAAAAGTTAA  
541 TGTTTTATCG GAAGCTGAGT TGACGTCAA AACAAAAGAA ACAAAGAAA AGTTAACCGC  
601 AACTTTTGAG CAGTTTAAAA AAGATACATT ACCAACAGAA CCAGAAAAA AGGTAGCAGA  
661 AGCTCAGAAG AAGGTTGAAG AAGCTAAGAA AAAAGCCGAG GATCAAAAAG AAAAAGATCG  
721 CCGTAACTAC CCAACCATA CTTACAAAAC GCTTGAACCT GAAATTGCTG AGTCCGATGT  
781 GGAAGTTAAA AAAGCGGAGC TTGAACTAGT AAAAGTGAAA GCTAAGGAAT CTCAAGACGA  
841 GGAAAAAATT AAGCAAGCAG AAGCGGAAGT TGAGAGTAAA CAAGCTGAGG CTACAAGGTT  
901 AAAAAAATC AAGACAGATC GTGAAGAAGC TAAACGAAAA GCAGATGCTA AGTTGAAGGA  
961 AGCTGTTGAA AAGAATGTAG CGACTTCAGA GCAAGATAAA CCAAAGAGGC  
GGGCAAAACG  
1021 AGGAGTTTCT GGAGAGCTAG CAACACCTGA TAAAAAGAA AATGATGCGA AGTCTTCAGA  
1081 TTCTAGCGTA GGTGAAGAAA CTCTTCCAAG CCCATCCCTT AATATGGCAA ATGAAAGTCA  
1141 GACAGAACAT AGGAAAGATG TCGATGAATA TATAAAAAA ATGTTGAGTG AGATCCAATT  
1201 AGATGGAAGA AAACATACCC CAAATGTCAA CTTAAACATA AAGTTGAGCG CAATTTAAAC  
1261 GAAGTATTTG TATGAATTA GTGTTTTAAA AGAGAACTCG AAAAAAGAAG AGTTGACGTC  
1321 AAAAACCAAA GCAGAGTTAA CCGCAGCTTT TGAGCAGTTT AAAAAAGATA CATTGAAACC  
1381 AGAAAAAAA GTAGCAGAAG CTGAGAAGAA GTTGAAGAA GCTAAGAAAA  
AAGCCAAGGA  
1441 TCAAAAAGAA GAAGATCGCC GTAACCTACC AACCAATACT TACAAAACGC TTGAACTTGA  
1501 AATTGCTGAG TCCGATGTGA AAGTTAAAGA AGCGGAGCTT GAAC TAGTAA  
AAGAGGAAGC  
1561 TAACGAATCT CGAAACGAGG AAAAAATTAA GCAAGCAAAA GAGAAAGTTG AGAGTAAAAA  
1621 AGCTGAGGCT ACAAGGTTAG AAAAAATCAA GACAGATCGT AAAAAAGCAG AAGAAGAAGC  
1681 TAAACGAAAA GCAGAAGAAT CTGAGAAAAA AGCTGCTGAA GCCAAACAAA AAGTGGATGC  
1741 TGAAGAATAT GCTCTTGAAG CTAAAATCGC TGAGTTGGAA TATGAAGTTC AGAGACTAGA  
1801 AAAGAGCTC AAAGAGATTG ATGAGTCTGA CTCAGAAGAT TATCTTAAAG AAGGCCTCCG  
1861 TGCTCCTCTT CAATCTAAAT TGGATACCAA AAAAGCTAAA CTATCAAAAC TTGAAGAGTT  
1921 GAGTGATAAG ATTGATGAGT TAGACGCTGA AATTGCAAAA CTTGAAGTTC AACTTAAAGA  
1981 TGCTGAAGGA AACAATAATG TAGAAGCCTA CTTTAAAGAA GGTTTAGAGA AAATACTGC  
2041 TGAGAAAAA GCTGAATTAG AAAAGCTGA AGCTGACCTT AAGAAAGCAG TTGATGAGCC  
2101 AGAAACTCCA GCTCCGGCTC CTCAACCAGC TCCGGCTCCA GAAAAACCAG  
CTGAAAAACC  
2161 AGCTCCAGCT CCAGCTCCAG AAAAACCAGC TCCAGCTCCA GAAAAACCAG  
CTCCAGCTCC  
2221 AGAAAAACCA GCTCCAGCTC CAGAAAAACC AGCTCCAGCT CCAGAAAAAC  
CAGCTCCAGC  
2281 TCCAGAAAAA CCAGCTCCAG CTCCAGAAAA ACCAGCTCCA GCTCCTAAAC CAGAAACTCC



47/50

*FIG.18B*

2341 AGAAACAGGC TGGAAACAAG AAAACGGTAT GTGGTACTTC TACAATACTG ATGGTTCAAT  
2401 GGCAACAGGC TGGCTCCAAA ACAATGGCTC ATGGTACTAC CTCAACAGCA  
ATGGCGTTAT  
2461 GGCGACAGGA TGGTTCCCAA ACAATGGTC  
//

48/50

FIG.19

LOCUS BG8090.DNA 1680 BP SS-DNA SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

CDS 220..1680

BASE COUNT 714 A 287 C 375 G 304 T 0 OTHER

ORIGIN

```

1 ATTGTATACG ACTCACTATA GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT
61 GGCCGCGGGA TTCGACGAAT AGCTGAAGAG GAAAAGCTAT TACATGAAGT TATAATCCCA
121 AATGGAAGCA TAAAGAGATA AATACAAAT TCGATTTATA TACAGTTCAT ATTGAAGTGA
181 TATAGTAAGG TTAAAGAAAA AATATAGAAG GAAATAAACA TGTTTGCATC AAAAAACGAA
241 AGAAAAGTAC ATTATTCAAT TCGTAAATTT AGTATTGGAG TAGCTAGTGT AGCTGTTGCC
301 AGTCTTTTTA TGGGAAGTGT GGTTCATGCG ACAGAGAAGG AGGTAAGTAC CCAAGTAGCC
361 ACTTCTTTTA ATAAGGCAAA TAAAAGTCAG ACAGAACATA TGAAAGCTGC TAAACAAGTC
421 GATGAATATA TAACAAAAAA GCTCCAATTA GATAGAAGAA AACATACCCA AAATGTCGGC
481 TTAATCACAAG AGTTGGGCGT AATTAAAACG GAGTATTTGC ATAGATTAAG TGTTTCAAAA
541 GAGAAGTCGG AAGCTGAGTT GCCGTCAGAA ATAAAAGCAA AGTTAGACGC AGCTTTTGAG
601 CAGTTTAAAA AAGATACATT ACCAACAGAA CCAGGAAAAA AGGTAGCAGA AGCTGAGAAG
661 AAGGTTGAAG AAGCTAAGAA AAAAGCCGAG GATCAAAAAG AAGAAGATCG TCGTAACTAC
721 CCAACCATTA CTTACAAAAC GCTTGAACCT GAAATTGCTG AGTCCGATGT GGAAGTTAAA
781 AAAGCGGAGC TTGAAGTAGT AAAAGAGGAA GCTAAGGGAT CTCGAAACGA GCAAAAAGTT
841 AACCAAGCAA AAGCGAAAAG TGAGAGTAAA CAAGCTGAGG CTACAAGGTT AAAAAAATC
901 AAGACAGATC GTGAACAAGC TGAGACTACA AGGTTAGAAA ACATCAAGAC AGATCGTGAA
961 AAAGCAGAAG AAGCTAAACG AAAAGCAGAT GCTAAAGAGC AAGATGAATC
AAAGAGGCGG
1021 GTAAAAGGAG GAGTTCCGGG AGAGCAAGCA ACACTTGATA AAAAAGAAAA
TGATGCGAAG
1081 TCTTCAGATT CTAGCGTAGG TGAAGAACT CTTCCAAGCC CATCCCTGAA ATCAGGAAAA
1141 AAGGTAGCAG AAGCTGAGAA GAAGGTTGCA GAAGCTGAGA AAAAAGCCAA
GGATCAAAAA
1201 GAAGAAGATC GCCGTAATA CCCAACCAAT ACTTACAAAA CGCTTGAAGT TGAAATTGCT
1261 GAGTCCGATG TGAAAGTTAA AGAAGCGGAG CTTGAAGTAG TAAAAGAGGA
AGCTAAGGAA
1321 TCTCGAAACG AGGAAAAAGT TAAGCAAGCA AAAGCGGAAG TTGAGAGTAA
AAAAGCTGAG
1381 GCTACAAGGT TAGAAAAAAT CAAGACAGAT CGTAAAAAAG CAGAAGAAGC TAAACGAAAA
1441 GCAGCAGAAG AAGATAAAGT TAAAGAAAAA CCAGCTGAAC AACCACAACC
AGCGCCGGCT
1501 CCTCAACCAG AAAAACCAGC TCCAGCTCCA AAACCAGAGA ATCCAGCTGA ACAACCAAAA
1561 GCAGAAAAAC CAGCTGATCA ACAAGCTGAA GAAGACTATG CTCGTAGATC AGAAGAAGAA
1621 TATAATCGCT TGAATCAACA GCAACCGCCA AAAACTGAAA AACCAGCACA ACCATCTACT
//

```



49/50

FIG.20

LOCUS L81905.DNA 1766 BP SS-DNA SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

CDS 217..1766

BASE COUNT 741 A 313 C 402 G 309 T 1 OTHER

ORIGIN -

1 GTATACGACT CACTATAGGG CGAATTGGGC CCGACGTCGC ATGCTCCCGG  
CCGCCATGGC  
61 CGCGGGATTC GACGAATAGC TGAAGAGGAA AAGCTATTAC ATGAAGTTAT AATCCCAAAT  
121 GGAAGCATAA AGAGATAAAT ACAAATTTCG ATTTATATAC AGTTCATATT GAAGTGATAT  
181 AGTAAGGTTA AAGAAAAAAT ATAGAAGGAA ATAAACATGT TTGCATCAAA AAGCGAAAGA  
241 AAAGTACATT ATTCAATTCG TAAATTTAGT GTTGGAGTAG CTAGTGTAGT TGTTGCCAGT  
301 CTTGTTATGG GAAGTGTGGT TCATGCGACA GAGAACGAGG GAGCTACCCA  
AGTACCCACT  
361 TCTTCTAATA GGGCAAATGA AAGTCAGGCA GAACAAGGAG AACAACCTAA AAAACTCGAT  
421 TCAGAACGAG ATAAGGCAAG GAAAGAGGTC GAGGAATATG TAAAAAAAT AGTGGGTGAG  
481 AGCTATGCAA AATCAACTAA AAAGCGACAT ACAATTACTG TAGCTCTAGT TAACGAGTTG  
541 AACAACATTA AGAACGAGTA TTTGAATAAA ATAGTTGAAT CAACCTCAGA AAGCCAACCTA  
601 CAGATACTGA TGATGGAGAG TCGATCAAAA GTAGATGAAG CTGTGTCTAA GTTTGAAAAG  
661 GACTCATCTT CTTCGTCAAG TTCAGACTCT TCCACTAAAC CGGAAGCTTC AGATACAGCG  
721 AAGCCAAACA AGCCGACAGA ACCAGGAGAA AAGGTAGCAG AAGCTAAGAA  
GAAGGTTGAA  
781 GAAGCTGAGA AAAAAGCCAA GGATCAAAAA GAAGAAGATC GTCGTAACCTA CCAACCATT  
841 ACTTACAAAA CGCTTGAACCT TGAAATTGCT GAGTCCGATG TGGAAGTTAA AAAAGCGGAG  
901 CTTGAACTAG TAAAGTGAA AGCTAACGAA CCTCGAGACG AGCAAAAAAT TAAGCAAGCA  
961 GAAGCGGAAG TTGAGAGTAA ACAAGCTGAG GCTACAAGGT TAAAAAAAT CAAGACAGAT  
1021 CGTGAAGAAG CAGAAGAAGA AGCTAACGA AGAGCAGATG CTAAAGAGCA  
AGGTAAACCA  
1081 AAGGGGCGGG CAAAACGAGG AGTTCCTGGA GAGCTAGCAA CACCTGATAA  
AAAAGAAAAT  
1141 GATGCGAAGT CTTGAGATTC TAGCGTAGGT GAAGAACTC TTCCAAGCCC ATCCCTGAAA  
1201 CCAGAAAAAA AGGTAGCAGA AGCTGAGAAG AAGGTTGAAG AAGCTAAGAA  
AAAAGCCGAG  
1261 GATCAAAAAG AAGAAGATCG CCGTAACTAC CCAACCAATA CTTACAAAAC GCTTGAACCT  
1321 GAAATTGCTG AGTCCGATGT GGAAGTTAAA AAAGCGGAGC TTGAACTAGT AAAAGAGGAA  
1381 GCTAAGGAAC CTCGAAACGA GGAAAAAGTT AAGCAAGCAA AAGCGGAAGT  
TGAGAGTAA  
1441 AAAGCTGAGG CTAAGAGTT AGAAAAAATC AAGACAGATC GTAAAAAAGC AGAAGAAGAA  
1501 GCTAAACGAA AAGCAGCAGA AGAAGATAAA GTTAAAGAAA AACCAGCTGA ACAACCACA  
1561 CCAGCGCCGG CTCCAAAAGC AGAAAAACCA GCTCCAGCTC CAAAACCAGA  
GAATCCAGCT  
1621 GAACAACCAA AAGCAGAAAA ACCAGCTGAT CAACAAGCTG AAGAAGAGTA TGCTCGTAGA  
1681 TCAGAAGAAG AATATAATCG CTTGACTCTA CAGCAACCGC CAAAACTGA AAAACCAGCA  
1741 CAACCATCTA CTCCAAAAAC AAANAC

//

50/50

## FIG. 21

LOCUS DBL6A.DNA 1590 BP SS-DNA SYN

DEFINITION -

ACCESSION -

KEYWORDS -

SOURCE -

FEATURES Location/Qualifiers

CDS 127..1590

BASE COUNT 701 A 261 C 339 G 288 T 1 OTHER

ORIGIN -

```
1 AAACATTAC ATGAAGTTAT AATCCCAAAT GGAAGCATAA AGAGATAAAT AAAAAATTCG
61 ATTTATATAC AGTTCATATT GAAGTGATAT AGTAAGGTTA AAGAAAAAAT ATAGAAGGAA
121 ATAATTATGT TTGCATCYAA AAGCGAAAGA AAAGTACATT ATTCAATTCG TAAATTTAGT
181 ATTGGAGTAG CTAGTGTAGC TGTGCTAGC TTGTTCTTAG GAGGAGTAGT CCATGCAGAA
241 GGGGTTAGAA GTGAGAATAC CCCCAAGGTT ACATCTAGTG GGGATGAAGT CGATGAATAT
301 AAAAAAAAAA TGTTGAGTGA GATCCAATTA GATAAAAGAA AACATACCCA CAATTTCCGC
361 TTAAACCTAA AGTTGAGCAG AATTAACACG GAGTATTTST ATAAATTAAA AGTTAATGTT
421 TTAGAAGAAA AGTCAAAAGC TGAGTTGACG TCAAAAACAA AAAAGAGAGT AGACGCAGCT
481 TTTGAGAAGT TAAAAAAGA TACATTGAAA CTAGGAGAAA AGGTAGCAGA AGCTCAGAAG
541 AAGGTTGAAG AAGCTAAGAA AAAAGCCAAG GATCAAAAAG AAGAAGATCA CCGTAAGTAC
601 CCAACCAATA CTTACAAAAC GCTTGAACCT GAAATTGCTG AGTCCGATGT GAAAGTTAAA
661 GAAGCGGAGC TTGAACCTAT GAAAGAGGAA GCTAAAACCT GAAACGAGGA CACAATTAAC
721 CAAGCAAAAG CGAAAGTTAA GAGTGAACAA GCTGAGGCTA CAAGGTTAAA AAAAATCAAG
781 ACAGATCGTG AACAAGCTGA GGCTACAAGG TTAGAAAACA TCAAGACAGA TCGTGAAAAA
841 GCAGAAGAAG CTAACGAAA AGCAGAAGCA GAAGAAGTTA AAGATAAACT AAAGAGGCGG
901 ACAAACGAG CAGTTCCTGG AGAGCCAGCA ACACCTGATA AAAAAGAAAA TGATGCGAAG
961 TCTTCAGATT CTAGCGTAGG TGAAGAACT CTTCCAAGCC CATCCCTGAA ATCAGGAAAA
1021 AAGGTAGCAG AAGCTCAGAA GAAGGTAGCA GAAGCTGAGA AAAAAGCCAA
GGATCAAAAA
1081 GAAGAAGATC GCCGTAAGTA CCCAACCAAT ACTTACAAAA CGCTTGACCT TGAAATTGCT
1141 GAGTCCGATG TGAAAGTTAA AGAAGCGGAG CTTGAACTAG TAAAGAGGA
AGCTAAGGAA
1201 TCTCGAAACG AGGAAAAAGT TAAGCAAGCA AAAGCGAAAG TTGAGAGTAA AAAAGCTGAG
1261 GCTACAAGGT TAGAAAAAAT CAAGACAGAT CGTAAAAAAG CAGAAGAAGC TAAACGAAGA
1321 GCAGCAGAAG AAGATAAAGT TAAAGAAAAA CCAGCTGAAC AACCACAACC
AGCGCCGGCT
1381 CCTCAACCAG AAAAACCAAC TGAAGAGCCT GAGAATCCAG CTCCAGCTCC
AAAACCTGAG
1441 AATCCAGCTG AACAACCAAA AGCAGAAAAA CCAGCTGATC AACAAGCTGA AGAAGACTAT
1501 GCTCGTAGAT CAGAAGAAGA ATATAATCGC TTGACTCAAC AGCAACCGCC AAAAAGTGA
1561 AAACCAGCAC AACCATCTAC TCCAAAAACA
```

//